Beiträge zu den Implikationen der digitalen Transformation für die Bildung

Kumulative Habilitation

zur Erlangung der Veni Legendi in

«Wirtschaftspädagogik»

an der

School of Management

Universität St.Gallen

eingereicht von

Juniorprofessor Dr. Josef Guggemos, MBR

Pädagogische Hochschule Schwäbisch Gmünd

Mentorin:

Professorin Dr. Sabine Seufert

Universität St.Gallen

Institut für Bildungsmanagement und Bildungstechnologien

Schwäbisch Gmünd 2022

Zusammenfassung

Die digitale Transformation hat weitreichende Implikationen für das Bildungs- und Beschäftigungssystem. In diesem Zusammenhang wird intensiv diskutiert, welche Kompetenzen Personen benötigen, um im beruflichen, gesellschaftlichen und privaten Kontext sachgerecht zu handeln. Lehrpersonen stehen bei der Förderung dieser Kompetenzen (neue) Bildungstechnologien zur Verfügung, die zunehmend künstliche Intelligenz nutzen. Dabei stellt sich die Frage, welche Kompetenzen Lehrpersonen zur Bewältigung dieser Aufgaben benötigen: zum einen, um zukunftsbefähigende Kompetenzen bei den Schüler:innen zu fördern und zum anderen, um Bildungstechnologien pädagogisch sinnvoll einzusetzen. Die vorliegende Habilitation adressiert diese Themenfelder und trägt damit zu einem besseren Verständnis der Implikationen der digitalen Transformation für die Bildung bei. Gegenstand der Habilitation sind zehn Studien, die in führenden englisch- und deutschsprachigen Zeitschriften bzw. Konferenzberichten veröffentlicht sind. Für die quantitativ-empirischen Beiträge dienen verschiedene Personengruppen als Datenbasis: Schüler:innen, Studierende, Lehrpersonen und Wirtschaftspädagog:innen. Dabei kam eine grosse Bandbreite an Methoden zum Einsatz, u.a. Strukturgleichungsmodelle, Item response theory, latente Profilanalysen und ausreisserrobuste Regressionen. Als Messinstrumente fungieren sowohl Selbsteinschätzungsskalen als auch Leistungstests. Die Habilitation leistet folgende Beiträge: Sie trägt zu einem besseren Verständnis des Konstrukts Computational Thinking bei – einer zentralen Kompetenz des 21. Jahrhunderts. Sie fördert das Verständnis für die gegenwärtig stark diskutierten Bildungstechnologien *Massive Open Online Courses* und *soziale Roboter*. Mit spezifischem Bezug zur Berufs- und Wirtschaftspädagogik legt sie konzeptionelle Grundlagen für eine Lernortkooperation mithilfe künstlicher Intelligenz. Im Forschungsfeld der nötigen Kompetenzen von Lehrpersonen im Kontext der digitalen Transformation unterstützt die Habilitation die Theorieentwicklung: Validiert wird ein konzeptionelles Modell. Dieses prognostiziert das Ausmass der Förderung technologiebezogener Kompetenzen bei den Schüler:innen und den Einsatz von Bildungstechnologien durch Lehrpersonen.

Inhalt

Studien

Studie 1: Computational thinking assessment – Towards more vivid interpretations
https://doi.org/10.1007/s10758-021-09587-2

Studie 2: On the predictors of computational thinking and its growth at the high-school level
https://doi.org/10.1016/j.compedu.2020.104060

Studie 3: Improving a MOOC to foster information literacy by means of a conjecture map
https://doi.org/10.1504/IJLT.2021.115470

Studie 4: Learners don't know best: Shedding light on the phenomenon of the K-12 MOOC in the context of information literacy
https://doi.org/10.1016/j.compedu.2022.104552

Studie 5: Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT
https://doi.org/10.1111/bjet.13006

Studie 6: Student acceptance of social robots in higher education – Evidence from a vignette study
https://files.eric.ed.gov/fulltext/ED626889.pdf

Studie 7: Neue Formen der Lernortkooperation mithilfe Künstlicher Intelligenz
https://doi.org/10.25162/9783515130752

Studie 8: Analyse beruflicher Tätigkeitsfelder von Wirtschaftspädagogen/-innen anhand von Daten des Karriereportals XING
https://doi.org/10.25162/zbw-2018-0019

Studie 9: Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends
https://doi.org/10.1016/j.chb.2020.106552

Studie 10: Teaching with and teaching about technology – Evidence for professional development of in-service teachers
https://doi.org/10.1016/j.chb.2020.106613

Hinweis: Mehrere Inhalte des nachfolgenden Kapitels sind ohne ausdrückliche Quellenangabe aus den Studien 1 bis 10 dieser Habilitation entnommen.

1 Einführung

Die digitale Transformation hat weitreichende Implikationen für das Bildungs- und Beschäftigungssystem. Bestehende Berufe ändern sich im hohen Tempo in ihren Anforderungsprofilen oder fallen ganz weg; gleichzeitig entstehen neue Berufe (Aepli et al., 2017; Brynjolfsson et al., 2018; Seufert, Guggemos, Tarantini & Schumann, 2019). Vor diesem Hintergrund gilt es, Kompetenzen zu fördern, die zur Bewältigung dieser Herausforderungen befähigen (Aepli et al., 2017; Becker & Spöttl, 2019). Generische Rahmenkonzepte existieren zu notwendigen Kompetenzen des 21. Jahrhunderts (van Laar et al., 2017). Rahmenkonzepte gibt es auch spezifisch zu den nötigen digitalen Kompetenzen (DigComp 2.2 der Europäischen Union, Vuorikari et al., 2022) und digitalen Kompetenzen am Arbeitsplatz (Oberländer et al., 2020). Diese Rahmenkonzepte weisen die Form von Kompetenzlisten auf. Dabei drängt sich die Frage auf, welche übergreifenden oder grundlegenden Fähigkeiten bei der Bewältigung des digitalen Wandels hilfreich sind. Im internationalen Kontext wird hier insbesondere das Konzept Computational Thinking (CT) diskutiert (Buitrago Flórez et al., 2017; Shute et al., 2017). Im deutschsprachigen Raum firmiert CT unter dem Begriff informatisches Denken (Bollin & Micheuz, 2019). Ziel von CT ist es, Methoden der Informatik domänenübergreifend in Problemlöseprozessen einzusetzen (Wing, 2006). An allgemeinbildenden Schulen wird CT zunehmend in Curricula integriert (Bocconi et al., 2016; Delcker & Ifenthaler, 2017). Die Lernenden erwerben dadurch potentiell Fähigkeiten, die in der beruflichen Bildung zur Lösung komplexer Problemstellungen genutzt werden können (Guggemos & Seufert, 2018). Auch in der kaufmännischen Domäne ist CT eine wichtige Ressource (Repenning et al., 2015). Die *Organisation kaufmännische Grundbildung Treuhand/Immobilien* schreibt im Kontext der Reform der kaufmännischen Grundbildung 2023 in der Schweiz (OKGT, 2022, S. 1): «Kaufleute von morgen handeln in agilen Arbeits- und Organisationsformen, interagieren in einem vernetzten Arbeitsumfeld und arbeiten mit neuen Technologien. Das setzt technische Fertigkeiten, Computational Thinking, Sozial- und Selbstkompetenzen sowie kritisches Denken und Kreativität voraus».

Neben Auswirkungen auf erforderliche Kompetenzen hat die digitale Transformation Implikationen für die Art des Unterrichtens. Die COVID-19 Pandemie hat die Bedeutung der Nutzung von Bildungstechnologien ins Bewusstsein einer breiten Öffentlichkeit gebracht (Delcker & Ifenthaler, 2021; Seufert et al., 2021). Die Forschung zum Einsatz von Bildungstechnologie blickt auf eine lange Tradition zurück (Scheiter, 2021). Ein Forschungsstrang sind Wirkungsstudien, bei denen versucht wird, die (kausalen) Effekte des Einsatzes bestimmter Technologien zu identifizieren (Seufert et al., 2021). In der beruflichen Bildung sind Beispiele die Untersuchung der Wirkung der Nutzung von Lernfabriken (Roll & Ifenthaler, 2021) oder interaktiver dynamischer Arbeitsblätter im Rechnungswesenunterricht (Berding et al., 2020). Insgesamt berichten Metaanalysen für den Einsatz von Bildungstechnologien mittlere aber nicht überwältigende Effekte (Hattie, 2009; Hillmayr et al., 2020; J-PAL, 2019; Tamim et al., 2011). Wirkungsstudien sind mit Problemen behaftet (Reeves & Lin, 2020; Scheiter, 2021). Sie ermöglichen keine

Antwort auf die Frage, ob zukünftig verfügbare Technologien zu höheren Effekten führen. Substantielle Fortschritte könnten von Bildungstechnologien zu erwarten sein, die künstliche Intelligenz nutzen (L. Chen et al., 2020). Die vorliegenden Metaanalysen zeigen zudem eine grosse Varianz in den erzielten Effekten. Es scheint daher fraglich, inwiefern Verallgemeinerungen möglich sind. Ein Beispiel ist der Einsatz von Tablets im Unterricht. Dem Einsatz von Tablets pauschal eine Wirkung zuzuschreiben dürfte verkürzt sein, da diese auf ganz unterschiedliche Weise eingesetzt werden können.

Neben Wirkungsstudien ist ein weiterer wichtiger Strang zur Beurteilung von Bildungstechnologien die Untersuchung der Technologieakzeptanz durch die Nutzenden. Wichtige Zielgruppen sind dabei Lehrpersonen und Schüler:innen (Ifenthaler & Schweinbenz, 2013; Raffaghelli et al., 2022; Scherer et al., 2019). Theorien zu diesem Zweck sind das Technologieakzeptanzmodell (TAM, Davis et al., 1989) und die Unified Theory of Acceptance and Use of Technology (UTAUT, Venkatesh et al., 2003). Sie zählen zu den am häufigsten genutzten Theorien bei der Untersuchung von Bildungstechnologien (Valtonen et al., 2022). Auch dieser Ansatz lässt sich kritisieren (Scheiter, 2021). Besonders der Rückgriff auf Selbsteinschätzungen zur Erfassung der Technologieakzeptanz-Konstrukte – mit bekannten Unzulänglichkeiten wie z.B. Selbstüberschätzung (Dunning et al., 2004) – sorgt für Kritik. Zweifel bestehen auch, inwiefern die Nutzungsintention für die tatsächliche Nutzung prädiktiv ist (Scherer et al., 2020). Ferner ist eine offene Frage, wie sich Wirkungsstudien und Technologieakzeptanzstudien verbinden lassen.

Die Digitalisierung ermöglicht neue Formen der Datengewinnung für quantitativ-empirische Forschung. Gerade im Bildungsbereich sind Datenzugänge mit hohen Hürden verbunden. Es sind Einwilligungen entsprechender Behörden und Schulen nötig; die datenschutzrechtlichen Anforderungen sind streng. Ferner ist die Bereitschaft zur Teilnahme an wissenschaftlichen Studien zunehmend gering. In der betriebswirtschaftlichen Forschung ist die automatisierte Gewinnung von online verfügbaren Daten bereits weit verbreitet (Claussen & Peukert, 2019). In der bildungswissenschaftlichen Forschung ist das eher nicht der Fall. Eine interessante Ausnahme bildet die Studie von Fütterer et al. (2021). Die Autor:innen konnten durch die Nutzung von Twitter-Daten schnell ein Bild zu Herausforderungen für Lehrpersonen während der Covid-19 Pandemie gewinnen.

Wie berichtet, weisen Metaanalysen zur Untersuchung der Wirkung des Einsatzes von Bildungstechnologien eine grosse Varianz in den Effekten auf. Vor diesem Hintergrund ist eine zunehmende Fokussierung auf Lehrpersonen und deren professionelle Kompetenzen zu sehen (Scheiter, 2021). Lehrpersonen setzen Bildungstechnologien unterschiedlich in Qualität und Quantität ein. Sie sind ein wichtiger Moderator für die Höhe der Effekte des Einsatzes von Bildungstechnologien (OECD, 2015). Neben der kompetenten Nutzung von Bildungstechnologien spielen Lehrpersonen auch bei der Förderung zukunftsbefähigender Kompetenzen bei den Schüler:innen eine wichtige Rolle. Die International Computer and Information Literacy Study 2018 (ICILS) fasst diese Doppelrolle zusammen mit (Fraillon et al., 2019, S. 175–176): «Teaching with and about ICT».

Übergreifend lassen sich drei wichtige Forschungsfelder im Kontext der digitalen Transformation mit Bezug auf die Bildung identifizieren. Abbildung 1 fasst diese drei Bereiche zusammen und zeigt wichtige Fragestellungen auf.

Digitalisierung als Lerngegenstand
Welche Kompetenzen benötigen Personen zur Bewältigung der digitalen Transformation?
- Wie lassen sich diese Kompetenzen operationalisieren?
- Wie lässt sich die Ausprägung dieser Kompetenzen prognostizieren?
- Wie lassen sich diese Kompetenzen fördern?

Digitalisierung als Methode
Wie kann Technologie Forschungs- und Bildungsprozesse unterstützen?
- Welche Lerneffekte sind zu erwarten?
- Wie ist die Technologieakzeptanz ausgeprägt?
- Wie lassen sich öffentlich verfügbare grosse Datensätze (big data) nutzen?

Kompetenzen von Lehrpersonen
- Welche Kompetenzen benötigen Lehrpersonen beim Unterrichten mit Technologie und über Technologie?
- Wie lassen sich diese Kompetenzen operationalisieren?
- Wie lassen sich diese Kompetenzen fördern?

Abbildung 1. Forschungsfelder und Fragestellungen zur digitalen Transformation in der Bildung.

Die vorliegende Habilitation adressiert alle drei Forschungsfelder. Kapitel 2 zeigt Forschungsdesiderate in den drei Bereichen auf. Kapitel 3 gibt einen Überblick über die zehn Studien der kumulativen Habilitation und ermöglicht eine Einschätzung deren Güte anhand geeigneter Kriterien. Kapitel 4 zeigt die methodische Bandbreite der Studien auf. Kapitel 5 beantwortet die in Kapitel 2 herausgearbeiteten Forschungsfragen. Kapitel 6 ist eine zusammenfassende Würdigung der Studien und verdeutlicht die Relevanz der Forschung für die Berufs- und Wirtschaftspädagogik. Ausserdem gibt Kapitel 6 einen Ausblick auf weiterführende eigene Forschung in den drei Bereichen.

2 Forschungsdesiderate

2.1 Computational Thinking als zentrale Kompetenz des 21. Jahrhunderts

CT, im deutschsprachigen Raum informatisches Denken, gilt als eine Schlüsselkompetenz des 21. Jahrhunderts (Voogt et al., 2015; Wing, 2006; Yadav et al., 2016). Computertechnologie durchdringt sämtliche Lebensbereiche und Arbeitsumgebungen (Barr et al., 2011; Buitrago Flórez et al., 2017; Wing, 2008). CT zielt darauf ab, diese Ressourcen domänenübergreifend zur Lösung von Problemen zu nutzen. Daher kann CT als eine weitere Kulturtechnik neben Lesen, Schreiben und Rechnen gelten (Barr et al., 2011). CT ist somit auch für den berufs- und wirtschaftspädagogischen Kontext von hoher Relevanz (Guggemos & Seufert, 2018). Beispielsweise können Entscheidungsprobleme automatisiert werden. Mithilfe von Simulationen können die Auswirkungen von Parameteränderungen, z.B. von Bezugspreisen, auf die Unternehmensziele abgeschätzt werden. Realistische Probleme, bei denen eine manuelle Berechnung schwierig ist, z.B. Berücksichtigung von Rabattstaffeln, können nun gelöst werden.

Als Konstrukt steckt CT noch in den Kinderschuhen (Shute et al., 2017). Eine wichtige Frage ist dabei, wie CT bei Schüler:innen gemessen werden kann (Tang et al., 2020; Weintrop et al., 2021). International sind vor allem zwei Instrumente etabliert (Shute et al., 2017): Der Computational Thinking test (CTt) (Román-González et al., 2017), ein Leistungstest, und die Computational Thinking Scales (CTS) (Korkmaz et al., 2017), ein Selbsteinschätzungsinstrument. Diese Instrumente übertrug ich ins Deutsche und validierte diese (Guggemos et al., 2019). Sie werden inzwischen auch von anderen Forschenden im deutschsprachigen Raum verwendet (Finke et al., 2022; Tsarava et al., 2022). Die Instrumente sind für den Einsatz in korrelativen Studien gut geeignet, beispielsweise um den Zusammenhang zwischen Geschlecht und CT zu untersuchen. Ein kritischer Punkt ist allerdings die Interpretation und Kommunikation der Testergebnisse.

Bezogen auf den CTt waren Aussagen möglich wie: 'Die Probandin hat 20 von 28 Aufgaben richtig beantwortet' oder 'Die Probandin gehört zu den 10% besten Testteilnehmenden'. Derartige Ergebnisse lassen sich schlecht kommunizieren und sind für ein formatives Assessment eher wenig nützlich. Internationale Vergleichsstudien, z.B. PISA oder ICILS, bilden daher Kompetenzniveaus (Fraillon et al., 2020; OECD, 2017). Ein Kompetenzniveau ist charakterisiert durch Aufgaben mit spezifischen Merkmalen, die eine Person systematisch lösen kann. Durch gezieltes Scaffolding (Kim & Hannafin, 2011) können Lernende dann auf ein höheres Kompetenzniveau gebracht werden. Beispielsweise löst eine Person auf Kompetenzniveau II Aufgaben systematisch richtig, die Funktionen enthalten, nicht aber Aufgaben mit Variablen. Eine Person auf Niveau II könnte durch eine gezieltes Scaffolding im Bereich des Konzepts der Variablen auf das zunächst nicht erreichbare Kompetenzniveau III gelangen. Der CTt bietet sich für Zwecke des Bildens von Kompetenzniveaus besonders gut an. Theoretisch fundierte (Brennan & Resnick, 2012) Aufgabenmerkmale wurden bereits bei der Entwicklung in ihrer Ausprägung dokumentiert. Eine analytische Verbindung (Hartig et al., 2012) zwischen den Aufgabenmerkmalen und den Aufgabenschwierigkeiten steht allerdings noch aus. Eine offene Frage ist vor diesem Hintergrund:
Welche Kompetenzniveaus lassen sich auf Basis des CTt bilden?

Der CTS erfasst die mehrdimensionale Struktur von CT, beispielsweise neben algorithmischem auch kritisches Denken und Kreativität. Dadurch ist allerdings ein übergreifendes CT-Profil für alle Schüler:innen unwahrscheinlich. Vielmehr sind unterschiedliche Profile zu erwarten. Schüler:innen könnten spezifische Stärken und Schwächen aufweisen, beispielsweise hohe Fähigkeiten im algorithmischen Denken in Verbindung mit einer geringen Kreativität. Pädagogisch ist das von Relevanz, weil Personen im selben Profil von ähnlichen Fördermassnahmen profitieren können (Hofmans et al., 2020). Zur Identifikation derartiger Gruppen bieten sich latente Profilanalysen an (Meyer & Morin, 2016; Spurk et al., 2020). Obwohl der Nutzen solcher personenzentrierter Methoden auch in der CT-Forschung betont wird (Román-González et al., 2019), gibt es bisher kaum Evidenz dazu. CTS bietet sich aufgrund der Mehrdimensionalität für diese Zwecke besonders an. Eine interessierende Forschungsfrage ist:

Welche latenten Profile lassen sich auf Basis des CTS identifizieren?

Um zu einem besseren Verständnis für ein Konstrukt beizutragen, ist die Prognose dessen Niveaus hilfreich (Newman et al., 2007). Aus pädagogischer Sicht wäre insbesondere auch Wissen zu den Prädiktoren des Lernzuwachses von CT nützlich. Ein umfassendes Verständnis für Variablen, die das Niveau und den Zuwachs von CT vorhersagen können, liegt nicht vor. Studien zur Vorhersage von CT adressieren häufig ausgewählte Bereiche, z.B. den familiären Hintergrund. Forschung zu Prädiktoren für CT ist hauptsächlich querschnittlich angelegt (z.B. Durak & Saritepeci, 2018); Studien zur Untersuchung von Prädiktoren der CT-Entwicklung fehlen. Ein weiteres Forschungsdesiderat ist die Untersuchung von CT Lernprozessen in naturalistischen Settings (Ilic et al., 2018; Lye & Koh, 2014); Befunde aus Wahlkursen sind aufgrund der Selbstselektion vermutlich nicht repräsentativ für CT Lernprozesse im Allgemeinen. Die Mehrzahl der vorliegenden Studien thematisiert zudem Grundschulen und die Sekundarstufe I. Das ist kritisch zu sehen, weil dadurch wichtige Phasen des CT Lernens unberücksichtigt bleiben (Tang et al., 2020). Es ergibt sich folgende Forschungsfrage:
Was sind die Prädiktoren für das Niveau und den Zuwachs von Computational Thinking auf Sekundarstufe II?

2.2 Digitalisierung als Methode

2.2.1 Massive Open Online Courses auf Sekundarstufe II

Im Hochschulkontext sind Massive Open Online Courses (MOOCs) ein vielbeachtetes Phänomen und Gegenstand intensiver Forschungsbemühungen (Joksimović et al., 2018; Reich & Ruipérez-Valiente, 2019; Zhu et al., 2020). Im Zuge dessen traten Probleme von MOOCs zutage, insbesondere die typischerweise sehr hohen Abbruchquoten (Hone & El Said, 2016). Durch die Einbindung von MOOCs in Präsenzlehrveranstaltungen (blended MOOCs) können die Vorteile von Präsenz- und Onlinelehre kombiniert werden (X. Chen et al., 2020; Zhu et al., 2020). Gerade in Form von blended MOOCs könnte diese Bildungstechnologie auf Sekundarstufe II vielversprechend sein: Hochwertige Lernarrangements werden zentral erstellt und können von einer Vielzahl an Lehrpersonen und Schüler:innen genutzt werden. Allerdings ist die Evidenz zu MOOCs im Schulbereich äusserst gering (Koutsakas et al., 2020). Im Vergleich zum Hochschulkontext mit über 19‘000 angebotenen Kursen[1], existiert lediglich eine verschwindend geringe Anzahl an MOOCs speziell für den Schulkontext. MOOCs aus dem Hochschulkontext im Schulkontext einzusetzen, scheint schwierig; hochwertige Lernarrangements berücksichtigen das Vorwissen und die Hintergründe der Lernenden (National Academies of Sciences, Engineering, and Medicine, 2018). Einem gestaltungsorientierten Forschungsparadigma folgend (McKenney & Reeves, 2014), wäre zunächst eine wichtige Forschungsfrage:
Wie kann ein hochwertiger MOOC für Schüler:innen auf Sekundarstufe II entwickelt werden?

[1] https://www.classcentral.com/report/mooc-stats-2021/

Die eigene Analyse der Literatur zeigt ein Defizit in den bisher verwendeten Methoden zur Evaluation von MOOCs im Schulkontext. Eine anerkannte Methode zur Evaluation von Bildungstechnologien ist die Analyse der damit verbundenen Lerneffekte (Scheiter, 2021). Das trifft auch auf die Evaluation von MOOCs zu (Reich, 2015). Eine wichtige Forschungsfrage ist:
Welcher Lerneffekt ergibt sich bei Einsatz eines MOOC bei Schüler:innen der Sekundarstufe II?

Im nächsten Schritt wäre dann zu klären, welche Variablen den Lerneffekt vorhersagen können: *Was sind die Prädiktoren für die Ausprägung des Lerneffekts beim Einsatz eines MOOC auf Sekundarstufe II?*

Ein weiterer anerkannter Ansatz zur Beurteilung von Bildungstechnologien ist die Untersuchung der Technologieakzeptanz der Lernenden (Raffaghelli et al., 2022). Dazu wird die Technologieakzeptanz (regelmässig die Nutzungsintention) mithilfe theoretisch begründeter Prädiktoren vorhergesagt. Es stellt sich die Frage:
Was sind die Prädiktoren für die Akzeptanz eines MOOC bei Schüler:innen auf Sekundarstufe II?

2.2.2 Soziale Roboter in der Hochschulbildung

Die Hochschulbildung ist ein Feld, mit dem sich die Wirtschaftspädagogik traditionell beschäftigt (Brahm et al., 2010; Euler, 2013; Gerholz & Sloane, 2011). Im Hochschulkontext ist es oftmals schwer möglich, den Studierenden individuelle Unterstützung zu bieten (Handke, 2018). Eine vielversprechende Möglichkeit kann in diesem Zusammenhang der Einsatz sozialer Roboter sein (Handke, 2018; Lehmann & Rossi, 2019). Soziale Roboter sind eine der fortschrittlichsten Bildungstechnologien; es lassen sich andere Technologien integrieren, z.B. intelligente tutorielle Systeme (Guggemos et al., 2022). Gegenüber virtuellen Agenten, z.B. Chatbots, haben soziale Roboter aufgrund ihrer physischen Präsenz Vorteile (Lehmann & Rossi, 2019; Li, 2015). Ihre Sensoren erlauben ihnen, eine Vielzahl an Daten zu sammeln und in Entscheidungsprozesse einzubeziehen. Soziale Roboter haben das Potential, ein wichtiger Bestanteil der Bildungsinfrastruktur zu werden (Belpaeme et al., 2018; Woo et al., 2021). Damit das Potential realisiert werden kann, sind allerdings wichtige Hürden zu nehmen. Eine davon ist die Technologieakzeptanz bei den Studierenden (Raffaghelli et al., 2022). Wenn sich die Studierenden weigern, soziale Roboter zu nutzen, wären diese nutzlos, unabhängig von einem tatsächlichen Wert. Elaborierte Studien zur Akzeptanz sozialer Roboter bei Studierenden existieren nicht. Zu klären wäre:
Was sind die Prädiktoren für die Technologieakzeptanz beim Einsatz sozialer Roboter im Hochschulkontext?

Die Antwort auf die vorhergehende Forschungsfrage ist ein erster Schritt zur Erschliessung der Technologieakzeptanz bei Studierenden. Soziale Roboter können allerdings ganz unterschiedliche Rollen im Hörsaal übernehmen. Beispiele sind Tutor oder Lehrassistent (Handke, 2020). Die Technologieakzeptanz könnte somit in Abhängigkeit von der ausgeübten Rolle des Roboters unterschiedlich ausfallen.

Ausserdem existieren vermutlich Studierendengruppen, die unterschiedliche Profile in ihrer Akzeptanz gegenüber sozialen Robotern aufweisen. Zwei wichtige Forschungsfragen sind damit:
Welche Rollen sozialer Roboter beim Einsatz im Hochschulkontext lassen sich empirisch trennen? Welche Profile der Akzeptanz sozialer Roboter zeigen sich bei Studierenden?

2.2.3 Spezifisch berufs- und wirtschaftspädagogische Fragestellungen

Ein Spezifikum der dualen beruflichen Bildung ist die Lernortkooperation (Aprea et al., 2020), d.h. das Zusammenwirken der Betriebe, der Berufsfachschulen und der überbetrieblichen Ausbildungsstätten zur optimalen Gesamtzielerreichung. Die Lernortkooperation gilt seit jeher als ein wichtiges Merkmal einer hohen Ausbildungsqualität (Wenner, 2018). Von neuen Technologien sind auch Impulse für die Lernortkooperation zu erwarten (Faßhauer, 2020). Entsprechende Konzepte wären zu erarbeiten und theoretisch zu fundieren. Auf konzeptioneller Ebene ist damit eine wichtige Forschungsfrage:
Welche (neuen) Möglichkeiten entstehen im Kontext der digitalen Transformation, Lernortkooperation, insbesondere durch künstliche Intelligenz, zu stärken?

Wirtschaftspädagogik ist ein polyvalenter Studiengang, angeboten in Deutschland, Österreich und der Schweiz; Absolvent:innen steht eine grosse Bandbreite an Berufen offen (Sektion BWP der DGfE, 2014). In welchen Tätigkeitfeldern Wirtschaftspädagog:innen tatsächlich agieren, ist allerdings eine schwierig zu beantwortende Frage (Sloane et al., 2004). Absolventenbefragungen liegen für einzelne Standorte vor, vornehmlich in Österreich (z.B. Stock et al., 2019). Ein repräsentatives Bild über Länder hinweg ist auf dieser Grundlage allerdings schwer möglich. Durch die Nutzung öffentlich verfügbarer grosser Datenmengen kann hier zur Erschliessung wirtschaftspädagogischer Tätigkeitsfelder ein wichtiger Beitrag geleistet werden. Aus forschungsmethodischer Perspektive wird dadurch gezeigt, dass *Digitalisierung als Methode* Implikationen über den Einsatz von Bildungstechnologie hinaus besitzt. Forschungsfragen können aufgrund neuer Wege der Datengewinnung besser als bisher beantwortet werden. Die konkrete Anwendung ist im vorliegenden Fall die Analyse wirtschaftspädagogischer Tätigkeitsfelder. Ein Beruf ist gekennzeichnet durch Struktur und Niveau (Bundesagentur für Arbeit, 2011). Die Fragestellungen hinsichtlich der beruflichen Tätigkeitsfelder von Wirtschaftspädagog:innen ist:
Wie ist die Berufsstruktur und das Berufsniveau bei Wirtschaftspädagog:innen ausgeprägt?

2.3 Kompetenzen von Lehrpersonen

Lehrpersonen spielen eine wichtige Rolle in Bildungsprozessen (Hattie, 2009). Die Ausprägung ihres professionellen Wissens kann einen bedeutenden Teil der Varianz in der Leistung zwischen Schulklassen erklären (Baumert et al., 2010). Im Kontext der digitalen Transformation sind die notwendigen professionellen Kompetenzen von Lehrpersonen ein Forschungsfeld mit zahlreichen offenen Fragen. Dieser Befund bewegte Sabine Seufert, Josef Guggemos und Michael Sailer zur Herausgabe eines Sonderhefts in der Zeitschrift *Computers in Human Behavior*. Wir identifizierten drei relevante Forschungsfelder, die acht international renommierte Autorenschaften bearbeiteten:

Welche Kompetenzen benötigen Lehrpersonen im Kontext der digitalen Transformation?
Wie können diese Kompetenzen gemessen werden?
Wie lassen sich diese Kompetenzen fördern?

Ein eigener Beitrag widmet sich dem ersten Forschungsfeld. Im Kontext der digitalen Transformation haben Lehrpersonen eine doppelte Rolle zu erfüllen (Fraillon et al., 2019). Zum einen bereiten sie die Schüler:innen auf die Anforderungen einer zunehmend digitalisierten Welt vor, zum anderen stehen ihnen neue Technologien zur Verfügung, mit denen sie Bildungsziele verfolgen können. Diese beiden Stränge wurden in der quantitativ-empirischen Literatur bisher getrennt diskutiert. Im Bereich der nötigen Kompetenzen zum Unterrichten mit Technologie fungiert das TPACK Rahmenkonzept (Koehler et al., 2013) häufig als theoretischer Hintergrund. Im Kontext der Förderung digitaler Kompetenzen bei den Schüler:innen werden andere Konzepte verwendet, z.B. *Teachers' emphasis on developing students' digital information and communication skills* (Siddiq et al., 2016) oder *Teaching in a Digital Environment Capacity* (Claro et al., 2018). Ein Forschungsdesiderat ist hier die Entwicklung eines konzeptionellen Modells, das die beiden Stränge in der Literatur kombiniert und sowohl die Nutzung digitaler Inhalte als auch digitaler Werkzeuge vorhersagen kann:
Welche Kompetenzen benötigen Lehrpersonen für die Nutzung von Bildungstechnologie?
Welche Kompetenzen benötigen Lehrpersonen, um digitale Kompetenzen bei ihren Schüler:innen zu fördern?

3 Überblick und Einordnung der Studien

Die in die Habilitation aufgenommenen zehn Studien beantworten die im vorhergehenden Kapitel aufgeworfenen Forschungsfragen. Sie haben grundsätzlich ein double-blind peer-review durchlaufen. Die Ausnahme ist Studie 5: Das *British Journal of Educational Technology* praktiziert ein single-blind peer-review. Im bildungswissenschaftlichen Kontext ist das Zeitschriftenranking des Verbands der Hochschullehrer für Betriebswirtschaft (VHB-Jourqual) nicht einschlägig. Somit dient zur Einordnung der englischsprachigen Beiträge der Journal Impact Factor 2021 (IF). Hervorzuheben sind die Studien 2, 4 und 9, die ich als Erst- bzw. Alleinautor in den Zeitschriften *Computers & Education* (IF = 11.2) sowie *Computers in Human Behavior* (IF = 9.0) veröffentlichte. Diese Zeitschriften gehören zu den renommiertesten im Bereich der Bildungstechnologien und Digitalisierung im Bildungskontext. Tabelle 1 zeigt die Publikationen und erlaubt eine Einschätzung deren Güte anhand geeigneter Kriterien.

Acht der zehn Beiträge sind in internationalen Formaten veröffentlicht, zwei Beiträge in *der Zeitschrift für Berufs- und Wirtschaftspädagogik*. Die *Zeitschrift für Berufs- und Wirtschaftspädagogik* ist die Zeitschrift mit der höchsten Relevanz und der dritthöchsten Reputation in der Berufs- und Wirtschaftspädagogik. Das ergab eine Umfrage in der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE) (Söll et al., 2014).

Zwei Beiträge der Habilitation sind Teil selbst herausgegebener Sonderhefte. Diese Mitherausgabe von Sonderheften belegt meine wissenschaftliche Aktivität in den entsprechenden Forschungsfeldern über die einzelnen Studien hinaus. In beiden Fällen fungierte ich als geschäftsführender Gastherausgeber. Der Peer-Review Prozess der eigenen Beiträge wurde von einer unabhängigen Person koordiniert.

Das Sonderheft *Technology-related knowledge, skills, and attitudes of pre- and in-service teachers* ist 2021 erschienen in der Zeitschrift *Computers in Human Behavior* (IF = 9.0). Herausgeber:innen sind Sabine Seufert, Josef Guggemos und Michael Sailer. Commissioning Editor war Paul A. Kirschner, einer der renommiertesten Forschenden im Bereich der Bildungswissenschaft. Er betreute auch die eigenen Beiträge zu diesem Sonderheft, beispielsweise die Koordination des Peer-Review Prozesses.

Das Sonderheft *Künstliche Intelligenz in der beruflichen Bildung: Zukunft der Arbeit und Bildung mit intelligenten Maschinen?!* ist 2021 erschienen in der *Zeitschrift für Berufs- und Wirtschaftspädagogik*. Herausgeber:innen sind Sabine Seufert, Josef Guggemos, Dirk Ifenthaler, Hubert Ertl und Jürgen Seifried. Das Sonderheft adressiert die Implikationen der künstlichen Intelligenz für die berufliche Bildung. Zum einen geht es darum, welche Auswirkungen die digitale Transformation auf verschiedene, z.B. kaufmännische, Berufe hat, zum anderen wie KI-basierte Technologien, z.B. Learning Analytics, Bildungsprozesse beeinflussen. Commissioning Editor war Jürgen Seifried.

Tabelle 1. Beiträge der kumulativen Habilitation und Einordnung deren Güte.

Nr.	**Digitalisierung als Lerngegenstand** - Computational Thinking als zentrale Kompetenz des 21. Jahrhunderts	
1	Guggemos, J., Seufert, S. & Román-González, M. (2022). Computational thinking assessment – Towards more vivid interpretations. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-021-09587-2	IF: 3.4
2	Guggemos, J. (2021). On the predictors of computational thinking and its growth at the high-school level. *Computers & Education*, 161, 104060. https://doi.org/10.1016/j.compedu.2020.104060	IF: 11.2
	Digitalisierung als Methode	
	Massive Open Online Courses auf Sekundarstufe II	
3	Moser, L., Guggemos, J. & Seufert, S. (2021). Improving a MOOC to foster information literacy by means of a conjecture map*. International Journal of Learning Technology*, 16(1), 65-86. https://doi.org/10.1504/IJLT.2021.115470	IF: 0.5
4	Guggemos, J., Moser, L. & Seufert, S. (2022). Learners don't know best: Shedding light on the phenomenon of the K-12 MOOC in the context of information literacy*. Computers & Education, 188, 104552.* https://doi.org/10.1016/j.compedu.2022.104552	IF: 11.2
	Soziale Roboter als fortgeschrittene Bildungstechnologie	
5	Guggemos, J., Seufert, S. & Sonderegger, S. (2020). Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology*, 51(5), 1864–1883. https://doi.org/10.1111/bjet.13006	IF: 5.3
6	Guggemos, J., Sonderegger, S. & Seufert, S. (2022). Student acceptance of social robots in higher education – Evidence from a vignette study. In *Proceedings of the 19th International Conference on Cognition and Exploratory Learning in Digital Age* (CELDA). https://files.eric.ed.gov/fulltext/ED626889.pdf	*

	Künstliche Intelligenz in der beruflichen Bildung	
7	Seufert, S., & Guggemos, J. (2021). Neue Formen der Lernortkooperation mithilfe Künstlicher Intelligenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik* (Beiheft 31), 183–214. https://doi.org/10.25162/9783515130752	**
	Nutzung online verfügbarer Daten - Berufliche Tätigkeitsfelder von Wirtschaftspädagog:innen	
8	Guggemos, J. (2018). Analyse beruflicher Tätigkeitsfelder von Wirtschaftspädagogen/-innen anhand von Daten des Karriereportals XING. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 114(4), 551–577. https://doi.org/10.25162/zbw-2018-0019	**
	Kompetenzen von Lehrpersonen	
9	Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 106552. https://doi.org/10.1016/j.chb.2020.106552	IF: 9.0
10	Guggemos, J., & Seufert, S. (2021). Teaching with and teaching about technology – Evidence for professional development of in-service teachers. *Computers in Human Behavior*, 115, 106613. https://doi.org/10.1016/j.chb.2020.106613	IF: 9.0

*Anmerkungen: IF = Journal Impact Factor im Jahr 2021. ** Der Beitrag unterlag einen double-blind peer-review und wurde mit dem Best Paper Award ausgezeichnet. ** Nach einer Umfrage in der Sektion Berufs- und Wirtschaftspädagogik die Zeitschrift mit der höchsten Relevanz und dritthöchsten Reputation (Söll et al., 2014, S. 518).

4 Empirische Methoden und Analyseverfahren

4.1 Übersicht

Tabelle 2 zeigt die Bandbreite der Studien der kumulativen Habilitation auf.

Tabelle 2. Übersicht zu Methode, Stichprobe, Messinstrument und primärem theoretischen Hintergrund der Studien.

Nr.	Methode	Stichprobe/ Messinstrumente	Primärer theoretischer Hintergrund
1	Item response theory (Rasch-Modellierung), Lineares logistisches Testmodell; Latente Profilanalyse	N = 202 Kantonsschüler:innen der 11. Jahrgangsstufe aus der Deutschschweiz/ Leistungstests und Selbsteinschätzungsinstrumente	CT Rahmenkonzept von Brennan und Resnick (2012); CT Rahmenkonzept der International Society for Technology in Education (ISTE, 2015)
2	Latentes Wachstumsmodell	Längsschnitt mit 3 Zeitpunkten: N_1 = 202, N_2 = 198, N_3 = 177 Kantonsschüler:innen der 11. Jahrgangsstufe aus der Deutschschweiz/ Leistungstests und Selbsteinschätzungsinstrumente	Modell zur Erklärung von Schulleistung der ICILS (Fraillon et al., 2019); Erwartungs-Valenz Modell der Leistungsmotivation (Wigfield et al., 2021); Selbstbestimmungstheorie (Ryan & Deci, 2000); Selbstkonzept (Shavelson et al., 1976)
3	Conjecture mapping; Interviews	N = 16 Lehrpersonen aus der Deutschschweiz	7i-Modell der Informationskompetenz (Seufert et al., 2020); Merkmale hochwertiger MOOCs (Egloffstein et al., 2019); conjecture mapping (Sandoval, 2014)
4	Strukturgleichungsmodellierung (CB-SEM); Change score und regressor variable model; Poweranalysen	N = 167 Schüler:innen auf Sekundarstufe II aus der Deutschschweiz, 74 davon an beruflichen Schulen/ Leistungstests und Selbsteinschätzungsinstrumente	Unified Theory of Acceptance and Use of technology (UTAUT) (Venkatesh et al., 2003) und deren Erweiterung UTAUT2 (Venkatesh et al., 2012); Erwartungs-Valenz Modell der Leistungsmotivation (Wigfield et al., 2021)

Nr.	Methode	Stichprobe/ Messinstrumente	Primärer theoretischer Hintergrund
5	Strukturgleichungsmodellierung (PLS-SEM); PLSpredict; Multigruppenanalyse; Importance-Performance Map Analyse	N = 462 Assessmentstudierende an der Universität St.Gallen/ Selbsteinschätzungsinstrumente	UTAUT (Venkatesh et al., 2003); Funktionen sozialer Robotern in der Hochschulbildung (Cooney & Leister, 2019); Merkmale hochwertiger Lernarrangements (Praetorius et al., 2018)
6	Konfirmatorische Faktorenanalyse; Latente Profilanalyse	N = 371 Assessmentstudierende an der Universität St.Gallen/ Selbsteinschätzungsinstrumente	Interactive, constructive, active, and passive (ICAP) Rahmenkonzept (Chi & Wylie, 2014); Rollen sozialer Roboter im Bildungskontext (Woo et al., 2021)
7	Konzeptioneller Beitrag	n/a	Digitales Ökosystem in der Bildung (Seufert, Guggemos & Moser, 2019); Modell integrativer Pädagogik (Tynjälä, 2008)
8	SMDM-Schätzer als ausreisserrobustes Verfahren; ausreisserrobuste MANOVA	N = 2'436 Profile von Wirtschaftspädagog:innen auf dem Karriereportal XING (via Webcrawling gewonnen)	Berufswahltheorie nach Gottfredson (Gottfredson, 1981) und Holland (Holland, 1997); Modell des Berufsprestiges (Ganzeboom & Treiman, 1996)
9	Konzeptioneller Beitrag	n/a	TPACK Rahmenkonzept (Koehler et al., 2013); Konzept der Orchestrierung von Lernprozessen (Dillenbourg, 2013); SQD Modell (Tondeur et al., 2012)
10	Strukturgleichungsmodellierung (PLS-SEM und CB-SEM); Multigruppenanalyse; Mediationsanalyse; Finite mixture segmentation	N = 212 Lehrpersonen an beruflichen Schulen in der Deutschschweiz/ Selbsteinschätzungsinstrumente	Skill, will, tool Modell (Knezek & Christensen, 2016); TPACK Rahmenkonzept (Koehler et al., 2013); DigComp 2.1 Rahmenkonzept (Carretero et al., 2017)

Anmerkung: Nr. = Nummer der Studie, siehe Tabelle 1.

4.2 Methoden

In den Beiträgen der kumulativen Habilitation nutze ich verschiedene Methoden. Der Fokus liegt auf quantitativ-empirischen Methoden, wobei auch qualitative Methoden (Interviews) und Methoden der gestaltungsorientierten Forschung (conjecture mapping) vertreten sind. Eine Zusammenfassung der verwendeten Methoden findet sich in Tabelle 2.

Insbesondere bei der Strukturgleichungsmodellierung gibt es eine kontroverse Debatte über die spezifischen Vor- und Nachteile kovarianzbasierter Strukturgleichungsmodellierung (CB-SEM) und partieller Kleinste-Quadrate Strukturgleichungsmodellierung (PLS-SEM) (Hair et al., 2019). Die Entscheidung für CB- oder PLS-SEM in der vorliegenden Habilitation basiert auf dem aktuellen Stand der Forschung (Sarstedt et al., 2022). Studie 6 nutzt PLS-SEM aufgrund der teilweisen Verwendung formativer Messmodelle. Ausserdem konnten mithilfe einer Importance-Performance Map Analyse tiefgreifende Erkenntnisse zur Technologieakzeptanz gewonnen werden (Ringle & Sarstedt, 2016). Studie 7 verwendet sowohl CB- als auch PLS-SEM. In Studie 7 sollte ein entwickeltes konzeptionelles Modell auf seine übergreifende Gültigkeit hin überprüft werden, was für CB-SEM sprach; zum anderen waren notwendige

Multigruppenvergleich aufgrund von Konvergenzproblemen im Fall von CB-SEM nur mit PLS-SEM möglich. Da, soweit vorliegend, die Ergebnisse von CB-SEM und PLS-SEM übereinstimmen, kann das als Evidenz für die Robustheit der Befunde gewertet werden (Sarstedt et al., 2016).

Zur Testvalidierung und Modellierung von CT dient die Item Response Theory (IRT) in Form der Rasch-Modellierung. Die IRT hat im Vergleich zur klassischen Testtheorie zahlreiche Vorteile (Hartig & Frey, 2013). Für die Kompetenzmodellierung besonders bedeutsam ist die Möglichkeit, Aufgabenschwierigkeiten und Personenfähigkeiten auf einer gemeinsamen Logit-Skala zu verorten. Das erlaubt, Kompetenzniveaus kriteriumsorientiert zu bilden. Ein Kompetenzniveau setzt sich aus Aufgaben mit bestimmten Eigenschaften zusammen, die die Proband:innen systematisch (z.B. zu mindestens 65%) richtig beantworten; Proband:innen auf einem geringeren Kompetenzniveau beantworten diese Aufgaben systematisch falsch (OECD, 2017).

Gerade bei der Gestaltung von Lernprozessen geht es um die personenbezogene Förderung. Hierzu können latente Profilanalysen hilfreich sein (Meyer & Morin, 2016; Spurk et al., 2020). Sie erlauben, spezifische Stärken- und Schwächenprofile bei den Proband:innen zu identifizieren (Morin & Marsh, 2015). Auf dieser Basis können dann die identifizierten Gruppen eine auf sie zugeschnittene Förderung erhalten. Gerade bei multidimensionalen Konstrukten wie beispielsweise CT oder der Akzeptanz sozialer Roboter sind heterogene Profile der Proband:innen wahrscheinlich.

Ausreisser können Schätzungen in quantitativ-empirischen Analysen verzerren (Begashaw & Yohannes, 2020). Bis auf eine Ausnahme scheint dieses Problem in den Studien dieser Habilitation wenig virulent. Die Proband:innen nahmen freiwillig an den Untersuchungen teil. Potentielle Ausreisser konnten leicht identifiziert und dann individuell beurteilt werden (Leys et al., 2019). Die Ausnahme ist der Datensatz in Studie 8 zur Untersuchung der beruflichen Tätigkeitsfelder von Wirtschaftspädagog:innen. Es handelt sich um Daten des Karriereportals XING. Hier zeigen deskriptive Analysen das Vorliegen einer Vielzahl an potentiellen Ausreissern. Die XING-Profile wurden teilweise von Nutzer:innen erkennbar über einen langen Zeitraum nicht gepflegt. Insgesamt lässt sich die Korrektheit der Angaben auf XING nur schwer validieren. Vor diesem Hintergrund kamen ausreisserrobuste statistische Verfahren zum Einsatz. Insbesondere diente ein SMDM-Schätzer bei den Regressionen für statistische Inferenzen (Koller & Stahel, 2011).

4.3 Messinstrumente

Quantitative Forschung ist auf die Operationalisierung von Konstrukten angewiesen. Die vorliegende Habilitation nutzt zwei Arten von Messinstrumenten: Leistungstests und Selbsteinschätzungen. Nicht selten wird die Ansicht vertreten, Selbsteinschätzung seien objektiven Leistungstests generell unterlegen (Conway & Lance, 2010). Die Validität eines Messinstruments sollte allerdings im Hinblick auf den intendierten Zweck bewertet werden (AERA et al., 2014). Im Kontext der Kompetenzen von Lehrpersonen zeigen Scherer et al. (2017) argumentativ fünf Vorteile von Selbsteinschätzungsinstrumenten auf: 1) Sie sind ein kostengünstiger, zuverlässiger und valider Indikator für die Selbstwirksamkeitserwartungen von Lehrpersonen; 2) Selbstwirksamkeitserwartungen sind wichtige Prädiktoren für die Absicht der Lehrpersonen, Technologie zu nutzen; 3) Selbsteinschätzungen sind auf zukünftiges Verhalten ausgerichtet; 4) Selbstwirksamkeitserwartungen korrelieren mit der Qualität des Unterrichts und positiven Bildungsergebnissen; 5) Selbstwirksamkeitserwartungen sind ein wünschenswertes Ergebnis der Aus- und Fortbildung bei Lehrpersonen. Zusammenfassend kann festgehalten werden, dass Selbsteinschätzungen und objektive Messungen unterschiedliche Konstrukte (Selbstwirksamkeitserwartungen vs. Leistung) erfassen. Beide Konstrukte haben ihre Berechtigung (Drummond & Sweeney, 2017).

Wenn es um die Erfassung von Lerneffekten geht, scheinen besonders Leistungstests geeignet. Studie 2 und 4 verwenden daher jeweils validierte Leistungstests (Guggemos et al., 2019; Seufert et al., 2020). Das ist insbesondere bei Studie 4 relevant. Studierende scheinen systematisch nicht in der Lage zu sein, ihre Informationskompetenz zuverlässig einzuschätzen (Mahmood, 2016).

4.4 Datengewinnung und Stichproben

Die kumulative Habilitation umfasst acht empirische Beiträge, die acht verschiedene Stichproben nutzen. Vier Personengruppen wurden betrachtet: Schüler:innen, Studierende, Lehrpersonen und Personen mit Wirtschaftspädagogikabschluss.

<u>Schüler:innen der Sekundarstufe II</u>

Daten für die Studien 1 und 2 gewann ich über das Netzwerk des Instituts für Bildungsmanagement und Bildungstechnologien (IBB) an drei Schulen im Kanton St.Gallen über drei Zeitpunkte von 2018 bis 2019 (N_1 = 202, N_2 = 198, N_3 = 173). Es handelt sich um Kantonsschüler:innen der 11. Jahrgangsstufe. Die 11. Jahrgangsstufe ist relevant, weil die Schüler:innen in diesem Schuljahr pro Woche 90 Minuten Informatikunterricht erhalten. Für Studie 1 dienen die Daten des Zeitpunkts 1 (N_1) als Stichprobe, für Studie 2 aller drei Zeitpunkte (N_1, N_2, N_3). Studie 4 nutzt Schüler:innendaten im Kontext eines selbst entwickelten MOOC (i-MOOC) mit 1‘032 Teilnehmenden im Jahr 2020. Die Teilnehmenden füllten auf freiwilliger Basis einen Fragebogen zum i-MOOC aus (N = 167).

<u>Assessmentstudierende der Universität St.Gallen</u>

Assessmentstudierende der Universität St.Gallen in der Veranstaltung Einführung in das wissenschaftliche Schreiben (Jahr 2019 [N = 476] und 2021 [N = 371]) fungieren als Datenbasis für Studie 5 und 6. Diese Personengruppe scheint besonders vielversprechend, weil im Hochschulkontext wenig Evidenz zum Einsatz sozialer Roboter vorliegt. Existierende Evidenz bezieht sich zudem regelmässig auf Studierende in den MINT-Fächern. Studierende aus den Sozialwissenschaften sind unterrepräsentiert. Ein weiterer Vorteil dieser Stichprobe ist, dass sich die damit gewonnenen Erkenntnisse potentiell gut auf

Schüler:innen der Sekundarstufe II übertragen lassen, weil die Proband:innen das Studium gerade erst aufgenommen haben.

<u>Lehrpersonen</u>

Als Teil der Evaluationsstrategie zur Verbesserung des i-MOOC im Design Research Prozess (Studie 3), wurden N = 16 Lehrpersonen interviewt, die mit ihren Klassen den i-MOOC im ersten Durchlauf (Jahr 2019) bearbeiteten.

Die Daten von N = 212 Lehrpersonen an Berufsfachschulen der Deutschschweiz im Jahr 2017 dienten zur Überprüfung des konzeptionellen Modells der notwendigen Kompetenzen im Kontext der digitalen Transformation. Dazu wurden alle Berufsfachschulen der Deutschschweiz kontaktiert. Neben fünf Partnerschulen nahmen sieben weitere Schulen teil. Die Schulleitungsteams der Schulen baten die Lehrpersonen, an der freiwilligen Onlineumfrage teilzunehmen.

<u>Wirtschaftspädagog:innen</u>

Gerade im bildungswissenschaftlichen Kontext ist es schwer, Datensätze ausreichender Grösse und Repräsentativität zu gewinnen. Webcrawling ist ein Ansatz, mit dem dieser Herausforderung begegnet werden kann. Bei Webcrawling werden Daten automatisiert aus öffentlich verfügbaren Onlinequellen gewonnen. Es ist mit zahlreichen Vorteilen verbunden (Landers et al., 2016): Es lassen sich sehr ökonomisch grosszahlige Datensätze über Regionen hinweg gewinnen. Die Daten sind zudem nicht durch das Verhalten der Forschenden beeinflusst (non-reaktiv). Trotz der Vorteile ist Webcrawling in der berufs- und wirtschaftspädagogischen Forschung bisher kaum eingesetzt worden. Für Studie 8 gewann ich die 2‘436 öffentlich verfügbaren Profile von Wirtschaftspädagog:innen auf dem Karriereportal XING. Zu diesem Zeitpunkt (November 2016) handelte es sich um das bei weitem mitgliederstärkste Karrierenetzwerk im deutschsprachigen Raum (XING AG, 2022). Auf dieser Grundlage scheint ein deutlich repräsentativeres Bild wirtschaftspädagogischer Tätigkeitsfelder möglich, als auf Basis von Absolventenbefragungen an einzelnen Standorten.

5 Beiträge der einzelnen Studien

5.1 Computational Thinking als zentrale Kompetenz des 21. Jahrhunderts

Basierend auf dem Computational Thinking test (CTt), lassen sich drei Kompetenzniveaus bilden (Studie 1). Diese sind gekennzeichnet durch spezifische Operationen, die Schüler:innen auf dieser Niveaustufe systematisch lösen können:

- Stufe I ist gekennzeichnet durch Aufgaben mit einer Ablaufstruktur ohne weitere Elemente wie Bedingungen (wenn, dann, sonst) oder Variablen. Ablaufstrukturen sind im kaufmännischen Kontext wichtige Operationen beispielsweise in der Reihenfolgeplanung oder im Projektmanagement.
- Stufe II umfasst Aufgaben, die Bedingungen und/oder Funktionen enthalten. Bedingungen und Funktionen sind CT-Kernelemente und in vielen Bereichen wichtig, z.B. für die effektive Nutzung

von Tabellenkalkulationssoftware wie Microsoft Excel. Eine solche Nutzung wiederum ist für eine Tätigkeit im Controlling von besonderer Bedeutung (Trachsel & Bitterli, 2020).

- Auf Stufe III enthalten die Aufgaben typischerweise Variablen, die ebenfalls ein CT-Kernkonzept darstellen. Variablen sind bei jeder Art von Entscheidungsproblem nützlich. Mithilfe von Variablen können Entscheidungsprobleme abstrakt modelliert werden. Am typischen Beispiel des Lieferantenvergleichs (Weber et al., 2014) wird das deutlich. Hier können die Parameter wie z.B. die Bezugspreise als Variablen deklariert werden. Bei einer Veränderung wäre statt des gesamten Entscheidungsproblems lediglich der Wert der entsprechenden Variable anzupassen.

Die Computational Thinking Scales (CTS) umfassen Selbsteinschätzungen zur Fähigkeit zum kritischen Denken, algorithmischem Denken, Kooperativität und Kreativität. Auf dieser Grundlage liessen sich vier latente Profile identifizieren (Studie 1):

- Profil 1 (kreative Personen mit Schwerpunkt auf Kooperation): Schüler:innen in diesem Profil schätzen ihr kreatives Denken als deutlich über dem neutralen Skalenmittelwert ein. Darüber hinaus erachten sie auch ihre Kooperativität und ihr kritisches Denken als über dem neutralen Skalenmittelwert. Ihre selbsteingeschätzte Fähigkeit zum algorithmischen Denken ist jedoch gering.
- Profil 2 (Computational Thinker auf niedrigem Niveau): Diese Schüler:innen schätzen sich selbst in allen CT-Dimensionen als unter dem Skalenmittelwert ein, d.h. sie halten sich in den vier CT-Dimensionen für nicht ausreichend leistungsfähig.
- Profil 3 (Computational Thinker mit geringer Kooperativität): Insgesamt schätzen die Schüler:innen ihr kreatives und kritisches Denken als sehr hoch ein. Ihr algorithmisches Denken liegt leicht über dem neutralen Skalenmittelwert. Allerdings berichten diese Schüler:innen eine sehr geringe Kooperativität.
- Profil 4 (Computational Thinker auf hohem Niveau): Diese Schüler:innen geben hohe Werte in allen vier CT-Dimensionen an.

Die Befunde aus Studie 1 haben wichtige pädagogische Implikationen. Die Kompetenzniveaus können eine personenspezifische Förderung unterstützen. Es wird deutlich, welche Aufgaben mit welchen Eigenschaften bestimmte Lernende noch nicht bewältigen können. Hier kann die Lehrperson gezielt ansetzen. Auf Grundlage der latenten Profile können Gruppen von Schüler:innen mit spezifischen Stärken und Schwächen identifiziert werden. Diese Information kann ebenfalls für die Instruktion im Klassenverbund genutzt werden. Schüler:innen aus unterschiedlichen Profilen könnten in Teams kombiniert werden, um im Lernprozess voneinander zu profitieren.

Studie 2 liefert Erkenntnisse zum CT-Niveau und dessen Zuwachs bei Schüler:innen auf Sekundarstufe II. Als abhängige Variable fungiert dabei der CTt als ein international anerkannter Leistungstest (Shute et al., 2017; Zhao & Shute, 2019). Eine erklärte Varianz von 70,4 % für das CT-Niveau und 61,2 % für das CT-Wachstum deutet auf einen guten Kompromiss zwischen Vollständigkeit und Sparsamkeit der Auswahl der Prädiktoren hin. Insgesamt spielt das Geschlecht eine wichtige Rolle bei der Erklärung des CT-Niveaus. Die Sprachfähigkeiten hingegen haben keinen Einfluss auf das Niveau oder die Entwicklung von CT. Da mathematische Fähigkeiten das CT-Niveau vorhersagen können, könnte CT enger mit mathematischen als mit sprachlichen Fähigkeiten zusammenhängen. Bezüglich der selbstberichteten Programmierfähigkeit ergibt sich ein signifikanter Zusammenhang mit dem CT-Niveau. Die Motivation in Form des CT-Selbstkonzepts und der selbstbestimmten Motivation spielt ebenfalls eine wichtige Rolle bei der Prognose des CT-Niveaus. Unterschiede in der Motivation könnten teilweise auch die Geschlechtsunterschiede im CT-Niveau erklären. Die Dauer der Computernutzung (PC und Laptop) steht in einem positiven Zusammenhang mit dem CT-Niveau und -Wachstum. Ferner sind die wahrgenommenen Lernmöglichkeiten während des untersuchten Schuljahres positiv prädiktiv für das CT-Wachstum. Nachfolgend beschriebene Implikationen für das Verständnis des Konstrukts CT ergeben sich.

Motivation ist ein wichtiger Prädiktor für das CT-Niveau. Bedeutsam scheint somit, für die Schüler:innen motivierende Instruktionsansätze zu wählen. Repenning (2017) argumentiert in diesem Zusammenhang, dass textbasierte Programmiersprachen die Schüler:innen demotivieren können und sich visuelle Programmiersprachen für die Förderung von CT anbieten. Im beruflichen Kontext könnten komplexe Lehr-Lernarrangements ein vielversprechender Ansatz sein (Achtenhagen, 2001), die die Schüler:innen mit Methoden der Informatik bearbeiten (Guggemos & Seufert, 2018).

In Zusammenschau mit anderen Studien scheint sich die Lücke zwischen der CT-Leistung männlicher und weiblicher Personen zum Vorteil männlicher Personen mit zunehmenden Alter nicht zu reduzieren (Román-González et al., 2017). Vielmehr weitet sie sich eher. Da CT eine zentrale Kompetenz des 21. Jahrhunderts und wichtig in Arbeitsprozessen ist (Yadav et al., 2018), gilt es dort anzusetzen. Wissen hierzu ist für Lehrpersonen in der beruflichen Bildung hilfreich: Schüler:innen treten mit unterschiedlichen CT-Niveaus die Ausbildung an und verfügen damit im unterschiedlichen Ausmass über eine wichtige Problemlöseressource, vergleichbar mit unterschiedlichen Mathematikfähigkeiten.

5.2 Digitalisierung als Methode

5.2.1 Massive Open Online Courses auf Sekundarstufe II

Informationskompetenz ist eine wichtige Kompetenz des 21. Jahrhunderts (Fraillon et al., 2019; Vuorikari et al., 2022). Dabei zeigen sich teils gravierende Defizite bei den Schüler:innen auf Sekundarstufe II (Seufert et al., 2020). Da Informationskompetenz für Schüler:innen aller Schularten in der Schweiz relevant ist, scheint ein MOOC zur Förderung dieser Kompetenz ein sinnvolles Unterfangen. Vor diesem Hintergrund entwickelten und verbesserten wir den i-MOOC zur Förderung von Informationskompetenz in einem Design Research Prozess (Studie 3). Tabelle 3 fasst zusammen, wie Kriterien eines guten MOOC (Egloffstein et al., 2019) bei der Entwicklung Berücksichtigung fanden.

Tabelle 3. Kriterien für einen hochwertigen MOOC (Egloffstein et al., 2019) und Umsetzung im i-MOOC.

Design Criteria	**i-MOOC Implementation**
A) Structure and clarity	
1. Learning goals *Learning goals are described comprehensively.*	▪ Detailed rubric of the learning goals on the i-MOOC website ▪ Main learning goals provided before each module in line with the 7i-framework for IL
2. Audience *The target audience is clearly described.*	▪ Clearly defined target group: Upper secondary-level students in German-speaking Switzerland
3. Requirements/effort *Course requirements are described sufficiently.*	▪ Available from the i-MOOC website ▪ No particular previous knowledge necessary ▪ Internet access, digital device
4. Course contents *The course contents are described in detail.*	▪ Detailed description in module 1 ▪ Detailed description on the i-MOOC website
5. Course structure *The course structure is clear.*	▪ Six modules structured in accordance with the 7i-framework ▪ Depicted in the preface of each module as an advance organizer
B) First principles of instruction (Merrill, 2002)	
6. Problem centeredness *The course tasks are linked to real-world problems.* *The course tasks are at the center of activities.*	▪ Overarching problem: Lena and Nico would like to become influencers ▪ All tasks geared toward solving the problem
7. Activation *The necessary prior knowledge is clearly described.* *The course elements (contents, tasks) build on prior knowledge.*	▪ Reflection questions to activate students' prior knowledge from personal experience ▪ The modules sequentially build on each other
8. Demonstration *New knowledge is being demonstrated in a coherent way.* *Media is being used adequately to demonstrate new knowledge.*	▪ Learning material in various forms: videos, images, and text ▪ Embedding the content in the overarching problem (Lena and Nico) and the experience of the learners
9. Application *New knowledge can be applied and practiced in a coherent way.* *The knowledge transfer to additional contexts is being promoted.*	▪ Quizzes in each module ▪ Transfer-tasks with peer feedback in each module ▪ Final overarching transfer module: Students have to create a video about how a self-selected profession may look in 2030
10. Integration *The reflection of new knowledge is being promoted.* *The discussion of new knowledge is being promoted.*	▪ Reflection consciously promoted as a facet of the 7i-framework ▪ Process and performance reflection after posttest ▪ Discussion of the acquired knowledge in the classroom
C) Additional principles of instruction	
11. Feedback *Feedback is an integral element of the course.*	▪ Two peer feedbacks for each transfer task ▪ Hints and model solutions for quizzes ▪ Visualization of students' progress and performance
12. Authentic resources *The course materials are authentic.*	▪ i-MOOC co-created by teachers and learners ▪ Lena and Nico are part of the target group
13. Differentiation *The course enables different learning pathways, according to learners' needs.*	▪ Transfer tasks allow for differentiation ▪ Teachers can supplement the i-MOOC with assignments and additional material
14. Cooperation/collaboration *The course promotes collaboration and cooperation.*	▪ Peer feedback tasks ▪ Classroom discussions ▪ Word-clouds and polls
15. Learner/activity orientation *The course promotes active learning.*	▪ Story-telling approach with peers (Nico and Lena) ▪ Students actively produce learning products: slides, videos ▪ Various task types

Quelle: Ergänzendes Onlinematerial zur Studie Guggemos et al. (2022).

Gestaltungsorientierte Forschung (Design Research) ist in der Berufs- und Wirtschaftspädagogik eine wichtige Strömung (Euler & Sloane, 2014; Tramm et al., 2017). Der vorliegende Beitrag zeigt auf, wie mithilfe einer Conjecture Map (Sandoval, 2014) ein MOOC entwickelt und verbessert werden kann. Hieran können sich andere Projekte orientieren und von den gesammelten Erfahrungen profitieren: bei der Entwicklung von MOOCs im Speziellen und von Bildungstechnologien im Allgemeinen.

Studie 4 evaluiert den i-MOOC mit quantitativer Methode. Der i-MOOC führt im Prä-Post- Vergleich zu einem Lerneffekt von d = 0.75, gemessen mit der Kurzversion eines validierten Leistungstests (Seufert et al., 2020). Das ist in der Zone akzeptabler Effekte (d > 0.4, Hattie, 2009). Signifikante Unterschiede im Lernzuwachs hinsichtlich des Geschlechts, Alters und Schultyps traten nicht auf. Insgesamt scheint es gelungen zu sein, einen für eine breite Schülerschaft angemessenen MOOC zu entwickeln.

Hinsichtlich der Technologieakzeptanz der Schüler:innen zeigt sich Erstaunliches: Von den einschlägigen Prädiktoren der UTAUT2 (Venkatesh et al., 2012) können nur hedonische Motive (wahrgenommener Spass und Freude bei der Bearbeitung) die Akzeptanz von MOOCs vorhersagen. Die Schüler:innen scheinen sich bei ihrer Entscheidung, ob die Lehrperson in Zukunft (weitere) MOOCs einsetzen soll, nahezu ausschliesslich von hedonischen Motiven leiten zu lassen. Das ist allerdings problematisch, denn hedonische Motive sind, wie Studie 4 zeigt, ein negativer Prädiktor für den Lernzuwachs. Neben diesem interessanten Befund demonstriert die Studie, wie Wirkungs- und Technologieakzeptanzstudie über die Erwartungs-Valenz Theorie der Leistungsmotivation kombiniert werden können.

Studie 4 hat zahlreiche Implikationen für den Einsatz von MOOCs im Schulkontext. Zunächst ist kritisch zu sehen, die Schüler:innen zu fragen, ob eine MOOC eingesetzt werden soll, wenn der Lernerfolg interessiert. Die Schüler:innen lassen sich bei der Frage nach dem Einsatz potentiell primär von hedonischen Motiven leiten. Zielführender wäre, zu fragen, ob ein bestimmter MOOC für sie einfach zu nutzen ist (Aufwandserwartung) und sie positive Lerneffekte (Leistungserwartung) sehen. Beides sind positive Prädiktoren für den Lernerfolg.

MOOCs sind ein vielbeachtetes Phänomen in der Hochschulbildung. Auch in der betrieblichen Bildung können sie eine bedeutende Rolle einnehmen (Egloffstein & Ifenthaler, 2017). Somit scheinen die in Studie 4 identifizierten Einstellungen der Schüler:innen gegenüber MOOCs problematisch. Wie beim Einsatz von Technologie allgemein, kann nicht erwartet werden, dass die Schüler:innen diese ohne entsprechende Instruktion zielorientiert nutzen können (Kirschner & Bruyckere, 2017). An Schulen könnte daher das Lernen mit MOOCs geübt werden. Das könnte die spätere Nutzung von MOOCs im hochschulischen und betrieblichen Kontext begünstigen.

MOOCs haben neben der Förderung des konkreten Lehrinhalts – beim i-MOOC Informationskompetenz – potentiell wünschenswerte Nebeneffekte. Die Schüler:innen können (mit Unterstützung der

Lehrpersonen) wichtige digitale Kompetenzen im Sinne des DigComp 2.2 Rahmenkonzepts der Europäischen Union erwerben (Vuorikari et al., 2022). Beispiele wären das Erstellen digitaler Inhalte und die digitale Kommunikation und Lernkompetenz mit digitalen Medien. Auch Lehrpersonen können durch die Nutzung von MOOCs ihre eigenen Kompetenzen steigern. Beispiele wären Kompetenzen im Bereich des Umgangs mit Daten und Learning Analytics (Berding et al., 2021; Ifenthaler et al., 2021).

5.2.2 Soziale Roboter in der Hochschulbildung

Ziel von Studie 5 ist es, die Akzeptanz sozialer Roboter für Lernzwecke bei Studierenden zu untersuchen. Ein Modell *Pepper* von SoftBank Robotics demonstrierte seine Fähigkeiten im Rahmen einer Vorlesung. Alle von der UTAUT implizierten Prädiktoren der Technologieakzeptanz – Leistungserwartung, Aufwandserwartung und sozialer Einfluss – sagen die Nutzungsintention positiv voraus. Die wahrgenommenen Eigenschaften des sozialen Roboters – Vertrauenswürdigkeit, Adaptivität, soziale Präsenz und Erscheinung – sagen die Nutzungsintention indirekt, mediiert durch die Leistungserwartung und die Aufwandserwartung, voraus. Überraschenderweise ist die Sorge der Studierenden um ihre Privatsphäre für die Nutzungsintention nicht negativ prädiktiv. Multigruppenanalysen und Finite-Mixture Segmentierungen belegen die Robustheit der Ergebnisse. Eine Importance-Performance Map Analyse zeigt die Adaptivität des Roboters als die Eigenschaft, die für die Erhöhung der Technologieakzeptanz besonders vielversprechend ist. Sie weist eine vergleichsweise geringe Performance und gleichzeitig eine relativ hohe Importance auf. Das heisst, der Zusammenhang von Adaptivität mit der Technologieakzeptanz ist im Vergleich zu den anderen Merkmalen des Roboters höher ausgeprägt und es gibt deutlichen Raum für Verbesserung im Bereich der Adaptivität aus der Sicht der Studierenden.

Studie 5 nach sind soziale Roboter auf dem derzeitigen Stand der Technik von den Studierenden eher nicht als Lernhilfe akzeptiert. Die Studie zeigt aber mögliche Stellschrauben auf. Insbesondere könnte es sich anbieten, die Adaptivität sozialer Roboter zu verbessern. Die Bedeutung von Adaptivität ist konform mit Forschung im pädagogischen Kontext (Brühwiler & Blatchford, 2011). Allerdings ist Adaptivität technisch gesehen schwierig zu steigern. Hierzu wäre insbesondere weitere Forschung auf dem Gebiet der künstlichen Intelligenz nötig. Interessanterweise sind zwar Bedenken in der Nutzung der eigenen Daten bei den Studierenden stark ausgeprägt, diese Bedenken können die Nutzungsintention allerdings nicht signifikant vorhersagen. Möglicherweise erklären lässt sich dieser Befund mit dem Privacy Paradox (Acquisti et al., 2015): Personen geben ihre Daten bereitwillig an Organisationen heraus und nutzen Dienste, denen sie nach eigenen Angaben misstrauen.

Studie 6 spezifiziert den Einsatz sozialer Roboter im Hochschulkontext über insgesamt acht Vignetten (Skilling & Stylianides, 2020). Die Vignetten präsentieren Rollen, die soziale Roboter übernehmen können. Sie wurden in einem Experteninterview validiert und offerieren den Studierenden eine angemessene Darstellung sozialer Roboter auf dem aktuellen Stand der Technik. Es handelt sich um soziale

Roboter als Präsentator, Lehrassistent, Anbieter formativen Assessments, Moderator, Tutor, Berater, Werkzeug (Unterrichtgegenstand) und Konversationspartner. Eine konfirmatorische Faktorenanalyse belegt die empirische Trennbarkeit der acht Rollen. Eine latente Profilanalyse ergab vier Profile, die eine sinnvolle Interpretation erlauben:

- Profil 1: Die Studierenden stehen allen acht Rollen sozialer Roboter in der Hochschullehre negativ gegenüber.
- Profil 2: Die Studierenden sind skeptisch gegenüber sozialen Robotern als Hilfsmittel bei der Instruktion. Sie bewerten jedoch den Einsatz sozialer Roboter als Werkzeug (Lerngegenstand) als positiv, z.B. zur Förderung von Computational Thinking.
- Profil 3: Die Studierenden haben eine differenzierte Sichtweise. Sie zeigen eine signifikant negative Akzeptanz gegenüber sozialen Robotern als Präsentator und Anbieter formativen Assessments, eine neutrale Einschätzung gegenüber sozialen Robotern als Moderator, Tutor und Konversationspartner und eine signifikant positive Einschätzung gegenüber sozialen Robotern als Lehrassistent, Berater und Werkzeug.
- Profil 4: Die Studierenden zeigen eine positive Einstellung zu allen acht Rollen sozialer Roboter.

Zusammenfassend stossen soziale Roboter als Lehrassistenten bei den Studierenden auf die höchste Akzeptanz. Soziale Roboter als Präsentator werden am negativsten bewertet; die Akzeptanz sozialer Roboter als Anbieter formativen Assessments ist ebenfalls negativ.

Studie 6 hat Implikationen für die Forschung zur Akzeptanz sozialer Roboter. Wie sich zeigt, kann nicht von einer allgemeinen Akzeptanz gesprochen werden. Diese ist vielmehr rollenabhängig. Zudem zeigt sich ein substantieller Anteil an Studierenden, die sozialen Robotern über alle Rollen hinweg negativ gegenüberstehen. Die (naive) Vermutung, Studierende seien fortgeschrittenen Technologien generell aufgeschlossen, lässt sich nicht halten. Die Befunde können helfen, bei der Entwicklung von Einsatzszenarien auf Rollen zu fokussieren, die eine hohe Akzeptanz aufweisen. Gleichzeitig sollte stets auf theoretisch begründbare Lernwirksamkeit geachtet werden (Sweller, 2020): Wie Studie 4 zeigt, ist die Technologieakzeptanz der Lernenden teilweise mit Vorsicht zu geniessen.

Zwar können soziale Roboter derzeit noch nicht flächendeckend über einen längeren Zeitraum hinweg eingesetzt werden. Als Forschungsgegenstand sind sie dennoch sehr interessant. Soziale Roboter geben künstlicher Intelligenz ein Gesicht. Darüber hinaus können hier zahlreiche andere Technologien integriert werden. Aufgrund der Vielzahl an Sensoren werden die moralischen Aspekte des Einsatzes künstlicher Intelligenz besonders deutlich. Hier eignen sich soziale Roboter gut, um ein Bewusstsein für dieses Thema zu schaffen und hierüber zu diskutieren. In der kaufmännischen Bildung könnten sie auch als Lerngegenstand eingesetzt werden. Beispielsweise könnte diskutiert werden, wie sich soziale Roboter als Verkaufsassistenten einsetzen lassen und Szenarien dazu entwickelt werden.

5.2.3 Spezifika der beruflichen Bildung

Studie 7 legt wichtige konzeptionelle Grundlagen für die Lernortkooperation mithilfe künstlicher Intelligenz. Zentral scheint, Lernortkooperation systematisch von den Lernenden her zu denken. Die Aktivitäten der Akteure der dualen beruflichen Bildung wären zu bündeln, um optimale Lernprozesse für die Lernenden zu ermöglichen. Hierzu ist Koordination, d.h. Abstimmung der Einzelaktivitäten nötig. Die gegenwärtige Ausprägung der Lernortkooperation kann als ein Gleichgewicht gesehen werden, in dem die Grenzkosten der Koordination dem Grenznutzen durch höhere Lernerfolge entsprechen. Anwendungen der künstlichen Intelligenz beeinflusst sowohl die Kosten- als auch die Nutzenseite. Zum einen verringert die fortgeschrittene Digitalisierung die Koordinationskosten. Zum anderen steigt der Druck, flexiblere, agilere und stärker digitalisierte Arbeits- und Lernformen in der Berufsbildung zu entwickeln, zu erproben und deren Qualität nachhaltig sicherzustellen. Mit anderen Worten, der Nutzen aus kooperativem Lernen steigt. Die beiden Mesosysteme Schule und Lehrbetrieb wären (besser) aufeinander abzustimmen. Digitale Ökosysteme (Seufert, Guggemos & Moser, 2019) sind hierfür eine wichtige Voraussetzung. Dieses Konzept eines digitalen Ökosystems in der Bildung wird vorgestellt. Auf der Mikroebene kann es sich anbieten, Einsatzszenarien zu entwickeln, um KI-basierte Anwendungen zu erproben. Dabei gilt es ethisch-moralische Herausforderungen des KI-Einsatzes zu berücksichtigen. Der Beitrag liefert konzeptionelle Grundlagen für eine technologieunterstütze Lernortkooperation, insbesondere unter Nutzung künstlicher Intelligenz. Auf dieses Fundament können weitere Beiträge zur Lernortkooperation aufbauen.

Studie 8 generiert umfassende Einblicke in berufliche Tätigkeitsfelder von Wirtschaftspädagog:innen ausserhalb des Schuldiensts. Die Berufsangaben wurden mit der Klassifikation der Berufe 2010 der Bundesagentur für Arbeit in Deutschland kodiert (Bundesagentur für Arbeit, 2011). Hinsichtlich der Struktur lassen sich mit 15 Berufsgruppen über 90% der Tätigkeiten von Wirtschaftspädagog:innen abdecken. Die beiden häufigsten sind mit über 40% Anteil *Personalwesen und -dienstleistung* sowie *Unternehmensorganisation und -strategie*. Hinsichtlich des Prestiges der ausgeübten Berufe (Niveau) zeigt sich ein positiver Einfluss der Höhe des akademischen Abschlusses. Besonders ausgeprägt ist der Prestigezuwachs bei einer Promotion im Vergleich zu allen anderen Abschlüssen. Evidenz dafür, dass Wirtschaftspädagoginnen systematisch Berufe auf geringerem Niveau ausüben als Wirtschaftspädagogen, konnte nicht gefunden werden. 16% der Wirtschaftspädagog:innen in der Stichprobe sind selbstständig tätig, am häufigsten als Unternehmensberater:in. Interessant ist, dass Wirtschaftspädagog:innen häufig in der Unternehmensberatung tätig sind. Dieses Tätigkeitsfeld fehlt im Basiscurriculum der Berufs- und Wirtschaftspädagogik (Sektion BWP der DGfE, 2014).

Studie 8 hat zahlreiche Implikationen. Nach Söll (2017) ist es wichtig, empirische Befunde in die Entwicklung von Studiengängen einzubeziehen. Die Ergebnisse aus Studie 8 können hierzu beitragen, weil ein umfassendes Bild von tatsächlichen Tätigkeitsfeldern von Wirtschaftspädagog:innen gezeichnet werden kann. Die im Basiscurriculum (Sektion BWP der DGfE, 2014) geforderten «praxisnahen Fragestellungen» können auf diese Weise empirisch fundiert werden. Studie 8 kann auch die Polyvalenz des Studiengangs Wirtschaftspädagogik demonstrieren: Es besteht eine grosse Bandbreite an attraktiven Optionen. Derartige Informationen können im Zuge einer Berufsberatung hilfreich sein. Methodisch erweitert Studie 8 das Spektrum der berufs- und wirtschaftspädagogischen Forschung. Es wird gezeigt, wie Daten automatisiert gewonnen und mit verschiedenen Datenbanken, beispielsweise dem etablierten O*Net (Peterson et al., 2001), verknüpft werden können. Dadurch hat der Beitrag insbesondere Implikationen für die Digitalisierung als (Forschungs-)Methode.

5.3 Kompetenzen von Lehrpersonen

Studie 9 ist ein konzeptioneller Überblicksartikel zum Stand der Forschung im Bereich der technologiebezogenen Kompetenzen von Lehrpersonen. Ausgehend davon werden gegenwärtige Strömungen aufgezeigt. Diese betreffen insbesondere Entwicklungen auf dem Gebiet der künstlichen Intelligenz (Zawacki-Richter et al., 2019). Es stellt sich die Frage, welche neuen oder veränderten Kompetenzen Lehrpersonen benötigen. Losgelöst vom bildungswissenschaftlichen Kontext zeichnet sich ab, dass künstliche Intelligenz besonders in der Zusammenarbeit mit Menschen Produktivitätsvorteile generieren kann (Dellermann et al., 2019). Burkhard et al. (2022) demonstrieren das konzeptionell für den Einsatz im Unterricht. Guggemos et al. (2022) zeigen anhand sozialer Roboter auf, wie eine solche Zusammenarbeit aussehen könnte. Lehrpersonen würden vor diesem Hintergrund in Zukunft Kompetenzen in der Zusammenarbeit mit künstlicher Intelligenz benötigen. Im Zuge dessen gibt es Bestrebungen, das TPACK Rahmenkonzept zu erweitern (Celik, 2022). Die Ausgestaltung der Zusammenarbeit von Mensch und Anwendungen künstlicher Intelligenz ist derzeit ein wichtiges Forschungsfeld mit zahlreichen offenen Fragen (Seeber et al., 2020). Einem gestaltungsorientierten Paradigma folgend, könnte es sich anbieten, die Denkrichtung anzupassen (Seufert et al., 2021): anstatt die nächsten Jahre auf der Grundlage der aktuellen Situation zu planen, Szenarien für die Zukunft zu entwerfen und aus dieser Perspektive mögliche Entwicklungsbereiche für das Lernen und Lehren der Schule der Zukunft zu explorieren. «The best way to predict the future is to invent it« (Kay, 1971).

Studie 10 liefert Beiträge zur Theorieentwicklung zu den notwendigen Kompetenzen bei Lehrpersonen und dem TPACK Rahmenkonzept. Präsentiert wird auf Grundlage des TPACK Rahmenkonzepts ein konzeptionelles Modell, das die notwendigen Kompetenzen zum Unterrichten mit digitalen Technologien und über digitale Inhalte integriert. Dadurch konnte ein im Vergleich zu bisher vorliegenden Arbeiten sparsameres Modell entwickelt werden. Ein solches sparsames Modell kann als ein wichtiger Beitrag zur Theorieentwicklung gelten (Whetten, 1989). Ein kritischer Punkt im TPACK Rahmenkonzept ist die Operationalisierung des Konstrukts des technologischen Wissens. Der Verweis auf konkrete Technologien führt zu einer schnell veralteten Definition (Koehler et al., 2013). Diese Herausforderung wurde gelöst, indem das DigComp 2.1 Rahmenkonzept der Europäischen Union (Carretero et al., 2017) zur

Spezifikation des technologischen Wissens dient. Wie die Versionsnummer 2.1 impliziert, ist das DigComp Rahmenkonzept dynamisch. Technologische Änderungen finden dadurch automatisch im entwickelten Rahmenkonzept Berücksichtigung. Durch die Erarbeitung neuer Items zur Erfassung der TPACK-Konstrukte konnte ein bestehendes Problem vorliegender Messinstrumente (z.B., Schmidt et al., 2009) behoben werden. Dieses liegt in der oftmals geringen diskriminanten Validität (Scherer et al., 2017), d.h. die Konstrukte lassen sich empirisch schwer trennen. Die entwickelten Instrumente weisen eine zufriedenstellende diskriminante Validität auf. Zudem fanden neuere Strömungen in der Forschung zum pädagogischen Wissen bei der Entwicklung Berücksichtigung (z.B. Voss et al., 2015), was die Inhaltsvalidität begünstigt. Kritisiert werden im TPACK Kontext auch die zu geringen Belege für eine prognostische Validität des TPACK Rahmenkonzepts (Graham, 2011). Hier konnte gezeigt werden, dass mithilfe des entwickelten konzeptionellen Modells auf Grundlage des TPACK Rahmenkonzepts die (selbstberichtete) Nutzung digitaler Technologien und der selbstberichtete Einsatz digitaler Inhalte im Unterricht prognostiziert werden kann.

6 Zusammenfassende Würdigung und Ausblick

6.1 Bedeutung der Forschung für die Wirtschaftspädagogik

Die vorliegende Habilitation enthält ausgewählte Studien aus meiner Forschungsaktivität mit Fokus auf international renommierte Zeitschriften. Die Studien haben gleichzeitig eine hohe Bedeutung für die Berufs- und Wirtschaftspädagogik. In allen Forschungsfelder der Habilitation präsentierte ich Beiträge auf der Sektionstagung Berufs- und Wirtschaftspädagogik der DGfE. Die berufs- und wirtschaftspädagogische Fachgesellschaft hat diese Beiträge dementsprechend im peer-review positiv beurteilt. Tabelle 4 zeigt die einschlägigen Konferenzbeiträge. Ferner hat die *Ständige Wissenschaftliche Kommission der Kultusministerkonferenz* in Deutschland in ihren Empfehlungen zur Digitalisierung im beruflichen Kontext diese Beiträge intensiv rezipiert (Köller et al., 2022).

Tabelle 4. Auf der Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE präsentierte Beiträge in den Forschungsfeldern der Habilitation.

Forschungsbereich	Vortrag
Computational Thinking	Guggemos, J. & Seufert, S. (2018). Computational Thinking – ein nützliches Konstrukt für die kaufmännische Bildung?! *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 3 - 5 in Frankfurt a.M.). Guggemos, J. & Seufert, S. (2021). Künstliche Intelligenz Literacy – ein wichtiges Konstrukt für die Berufs- und Wirtschaftspädagogik?! *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 15 - 17 online).
Massive Open Online Courses auf Sekundarstufe II	Seufert, S., Guggemos, J., Moser, L., & Sonderegger, S. (2019). Entwicklung und Validierung eines MOOC zur Förderung von Informationskompetenz. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 25 - 27 in Graz). Guggemos, J., Moser, L., & Seufert, S. (2021). Massive Open Online Courses auf Sekundarstufe II – Akzeptanz durch Lernende. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 15 - 17 online).

Forschungsbereich	Vortrag
Soziale Roboter	Guggemos, J., Seufert, S., & Sonderegger, S. (2020). Akzeptanz sozialer Roboter als Lernhilfe in der Hochschulbildung. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 9 - 10 online). Guggemos, J., Sonderegger, S., & Seufert, S. (2022). Einsatz sozialer Roboter im Hochschulkontext: Ergebnisse einer Studie zur Technologieakzeptanz. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 26 - 28 in Freiburg).
Spezifika der Berufs- und Wirtschaftspädagogik	Guggemos, J. (2017). Analyse beruflicher Tätigkeitsfelder von Wirtschaftspädagogen mit Daten des Karriereportals XING. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 25 - 27 in Stuttgart). Seufert, S. & Guggemos, J. (2018). Flexibilisierung der Berufsbildung in der Schweiz im Kontext fortschreitender Digitalisierung – eine explorative Studie. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 3 - 5 in Frankfurt a.M.). Seufert, S. & Guggemos, J. (2019). Lernortkooperation in der Schweiz: Gestaltung eines digitalen Ökosystems für die berufliche Bildung? *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 25 - 27 in Graz).
Kompetenzen von Lehrpersonen	Seufert, S., Guggemos, J., & Tarantini, E. (2018). Digitale Kompetenzen von berufsbildenden Lehrpersonen – Ergebnisse einer empirischen Studie. *Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE* (September 3 - 5 in Frankfurt a.M.).

Anmerkung: Präsentierende Person unterstrichen.

Abbildung 2 zeigt die Zuordnung der zehn Studien zu den in der Einführung aufgezeigten Forschungsfeldern zur digitalen Transformation in der Bildung. In allen drei Bereichen konnte eine Veröffentlichung in einer der Top-Zeitschriften des Feldes in Erstautorenschaft realisiert werden (*Computers & Education* und *Computers in Human Behavior*). Das belegt die internationale Anschlussfähigkeit der eigenen Forschung.

Digitalisierung als Lerngegenstand
Welche Kompetenzen benötigen Personen zur Bewältigung der digitalen Transformation?

- Computational Thinking als eine zentrale Kompetenz des 21. Jahrhunderts (Studie 1 und 2)

Digitalisierung als Methode
Wie kann Technologie Forschungs- und Bildungsprozesse unterstützen?

- MOOCs (Studie 3 und 4)
- Soziale Roboter (Studie 5 und 6)
- Lernortkooperation (Studie 7)
- Nutzung von Daten des Karriereportals XING (Studie 8)

Kompetenzen von Lehrpersonen (Studie 9 und 10)

Abbildung 2. Abdeckung der drei Forschungsfelder zur digitalen Transformation in der Bildung.

6.2 Ausblick

Die in der Habilitation adressierten drei Felder bieten inhaltlich und methodisch die Grundlage für weitere Forschung. Als Blaupause können dafür die Studien 3 und 4 zu MOOCs dienen: In einem Design Research Prozess werden hochwertige technologieunterstützte Lernszenarien entwickelt, die dann mit quantitativ-empirischer Methode evaluiert werden.

Digitalisierung als Lerngegenstand

Wie dargelegt, spielt künstliche Intelligenz eine zunehmend wichtige Rolle im Bildungsbereich. Das gilt nicht nur für die Bildungstechnologien, sondern auch für die Kompetenzen bei den Schüler:innen. Deutlich wird das beispielsweise an der Anpassung des DigComp Rahmenkonzepts der Europäischen Union von der Version 2.1 auf 2.2. Hier sind nun explizit Kompetenzen im Bereich künstlicher Intelligenz enthalten (Vuorikari et al., 2022). Im internationalen Kontext wird dabei das Konzept der *Artificial Intelligence Literacy* diskutiert (Long & Magerko, 2020; Ng et al., 2021). Diese Konzept brachte ich in die berufs- und wirtschaftspädagogische Diskussion ein (Guggemos & Seufert, 2021). Computational Thinking bietet sich als Grundlage zur Vermittlung von Kompetenzen im Bereich künstlicher Intelligenz an (Wong et al., 2020). Untersucht werden soll, wie Computational Thinking zu einer effizienten Lösung von Problemen beitragen kann, mit denen kaufmännische Auszubildende typischerweise konfrontiert sind. Das inkludiert die Frage, wie Anwendungen der künstlichen Intelligenz diese Problemlöseprozesse unterstützen können. Wie eigene Vorarbeiten zeigen, treten zahlreiche ethische Herausforderungen beim Einsatz künstlicher Intelligenz auf (Jobin et al., 2019). Wie Nutzer:innen diesen Herausforderungen begegnen können, stellt ein interessantes Forschungsfeld dar.

Digitalisierung als Methode

Forschung im Bereich sozialer Roboter ist vielversprechend, weil sie eine Vielzahl an weiteren Bildungstechnologien integrieren lassen, beispielsweise Learning Management Systeme (Sonderegger, 2022). In meiner Forschung zu sozialen Robotern plane ich, in Zukunft verstärkt mit Experimentaldesigns zu arbeiten: Auf Grundlage pädagogisch-psychologischer Theorien werden sinnvolle Einsatzszenarien für soziale Roboter identifiziert und in einem Design Research Prozess entwickelt und abschliessend in einem Experimentaldesign evaluiert. Interessierende abhängige Variablen sind Lernzuwächse und die Technologieakzeptanz, inklusive moralischen Bedenken des Einsatzes. Wie in Studie 4 aufgezeigt, scheint die Kombination aus Wirkungs- und Technologieakzeptanzstudie vielversprechend.

Kompetenzen von Lehrpersonen

Lehrpersonen und ihr professionelles Wissen spielen eine zentrale Rolle bei der Bewältigung der digitalen Transformation. Simulationen sind ein sehr effektives Instrument, um Kompetenzen zu fördern (Chernikova et al., 2020). Hier wäre interessant, Simulationen – auch unter Nutzung künstlicher Intelligenz – zu entwickeln und anschliessend in ihrer Wirkung zu überprüfen. Derartige Simulationen könnten auch genutzt werden, um technologiebezogene Kompetenzen von Lehrpersonen zu überprüfen. Das wäre eine Möglichkeit, den in der Literatur häufig kritisierten Rückgriff auf Selbsteinschätzungsinstrumente zu vermeiden (Scheiter, 2021). Lehrpersonen benötigen ferner (in Zukunft) Kompetenzen im Umgang mit künstlicher Intelligenz (Celik, 2022). Diese Kompetenzfacetten für die berufliche Bildung systematisch auszudifferenzieren, ist ein interessantes Forschungsfeld.

Literatur

Achtenhagen, F. (2001). Criteria for the development of complex teaching-learning environments. *Instructional Science*, *29*(4/5), 361–380. https://doi.org/10.1023/A:1011956117397

Acquisti, A., Brandimarte, L. & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Aepli, M., Angst, V., Iten, R., Kaiser, H., Luthi, I. & Schweri, J. (2017). *Die Entwicklung der Kompetenzanforderungen auf dem Arbeitsmarkt im Zuge der Digitalisierung*. Staatssekretariat für Wirtschaft.

AERA, APA & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://eric.ed.gov/?id=ED565876

Aprea, C., Sappa, V. & Tenberg, R. (2020). Konnektivität und integrative Kompetenzentwicklung. Einleitung zum Themenheft. *Zeitschrift für Berufs- und Wirtschaftspädagogik*(Beiheft 29), 9–13.

Barr, D., Harrison, J. & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology*, *38*(6), 20–23. https://eric.ed.gov/?id=EJ918910

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Becker, M. & Spöttl, G. (2019). Auswirkungen der Digitalisierung auf die berufliche Bildung am Beispiel der Metall- und Elektroindustrie. *Zeitschrift für Erziehungswissenschaft*, *22*(3), 567–592. https://doi.org/10.1007/s11618-019-00869-1

Begashaw, G. B. & Yohannes, Y. B. (2020). Review of Outlier Detection and Identifying Using Robust Regression Model. *International Journal of Systems Science and Applied Mathematics*, *5*(1), 4. https://doi.org/10.11648/j.ijssam.20200501.12

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), 1-9. https://doi.org/10.1126/scirobotics.aat5954

Berding, F., Jahncke, H. & Albersmann, M. (2020). Dynamische und interaktive Arbeitsblätter im Rechnungswesenunterricht. Eine Wirkungsstudie im Kontrollgruppendesign. In F. Berding, H. Jahncke & A. Slopinski (Hrsg.), *Moderner Rechnungswesenunterricht 2020* (S. 277–317). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-31146-9_13

Berding, F., Jahncke, H. & Holt, K. (2021). Learning Analytics in der Wirtschaftspädagogik: Eine Simulationsstudie für die Anwendung überwachten maschinellen Lernens für Inhaltsanalysen am Beispiel von Grundvorstellungen und (Selbst-)Reflexionskompetenz. In S. Seufert, J. Guggemos, D. Ifenthaler, H. Ertl & J. Seifried (Hrsg.), *Beiheft 31 der Zeitschrift für Berufs- und Wirtschaftspädagogik. Künstliche Intelligenz in der beruflichen Bildung: Zukunft der Arbeit und Bildung mit intelligenten Maschinen?!* (S. 237–292). Steiner.

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A. & Engelhardt, K. (2016). *Developing computational thinking in compulsory education*. Publications Office of the European Union. https://doi.org/10.2791/792158

Bollin, A. & Micheuz, P. (2019). Computational Thinking on the Way to a Cultural Technique. In D. Passey, R. Bottino, C. Lewin & E. Sanchez (Hrsg.), *IFIP Advances in Information and Communication Technology. Empowering Learners for Life in the Digital Age* (Bd. 524, S. 3–13). Springer International Publishing. https://doi.org/10.1007/978-3-030-23513-0_1

Brahm, T., Jenert, T. & Meier, C. (2010). *Hochschulentwicklung als Gestaltung von Lehr- und Lernkultur: Eine institutionsweite Herangehensweise an lehrbezogene Veränderungsprojekte an Hochschulen*. IWP Arbeitsberichte.

Brennan, K. & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *American Educational Research Association meeting, Vancouver, BC, Canada*, 1–25. https://www.media.mit.edu/publications/new-frameworks-for-studying-and-assessing-the-development-of-computational-thinking/

Brühwiler, C. & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, *21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Brynjolfsson, E., McAfee, A. & Henzler, H. A. (2018). *The Second Machine Age: Wie die nächste digitale Revolution unser aller Leben verändern wird*. Plassen Verlag.

Buitrago Flórez, F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S. & Danies, G. (2017). Changing a generation's way of thinking: Teaching computational thinking through programming. *Review of Educational Research*, *87*(4), 834–860. https://doi.org/10.3102/0034654317710096

Bundesagentur für Arbeit. (2011). *Klassifikation der Berufe 2010: Band 1: Systematischer und alphabetischer Teil mit Erläuterungen*. https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB-2010/Generische-Publikationen/KldB2010-Printversion-Band1.pdf

Burkhard, M., Guggemos, J. & Seufert, S. (2022). Who is Best Suited for the Job? Task Allocation Process Between Teachers and Smart Machines Based on Comparative Strengths. In B. Csapó & J. Uhomoibhi (Hrsg.), *Communications in Computer and Information Science. Computer Supported Education* (S. 3–23). Springer International Publishing. https://doi.org/10.1007/978-3-031-14756-2_1

Carretero, S., Vuorikari, R. & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. European Union (EU). https://doi.org/10.2760/38842

Celik, I. (2022). Towards Intelligent-TPACK: An empirical study on teachers’ professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Chen, L., Chen, P. & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, X., Zou, D., Xie, H. & Di Zou (2020). Fifty years of British Journal of Educational Technology: A topic modeling based bibliometric perspective. *British Journal of Educational Technology*, *51*(3), 692–708. https://doi.org/10.1111/bjet.12907

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T. & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, *90*(4), 499–541. https://doi.org/10.3102/0034654320933544

Chi, M. T. H. & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Claro, M., Salinas, A., Cabello-Hutt, T., San Martín, E., Preiss, D. D., Valenzuela, S. & Jara, I. (2018). Teaching in a Digital Environment (TIDE): Defining and measuring teachers' capacity to develop students' digital information and communication skills. *Computers & Education*, *121*, 162–174. https://doi.org/10.1016/j.compedu.2018.03.001

Claussen, J. & Peukert, C. (2019). Obtaining Data from the Internet: A Guide to Data Crawling in Management Research. *SSRN Electronic Journal.* Vorab-Onlinepublikation. https://doi.org/10.2139/ssrn.3403799

Conway, J. M. & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*, *25*(3), 325–334. https://doi.org/10.1007/s10869-010-9181-6

Cooney, M. & Leister, W. (2019). Using the Engagement Profile to design an engaging robotic teaching assistant for students. *Robotics*, *8*(1), 1–26. https://doi.org/10.3390/robotics8010021

Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Delcker, J. & Ifenthaler, D. (2017). Computational Thinking as an Interdisciplinary Approach to Computer Science School Curricula: A German Perspective. In P. J. Rich & C. B. Hodges (Hrsg.), *Emerging Research, Practice, and Policy on Computational Thinking* (S. 49–62). Springer International Publishing.

Delcker, J. & Ifenthaler, D. (2021). Teachers' perspective on school development at German vocational schools during the Covid-19 pandemic. *Technology, Pedagogy and Education*, *30*(1), 125–139. https://doi.org/10.1080/1475939X.2020.1857826

Dellermann, D., Ebel, P., Söllner, M. & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492. https://doi.org/10.1016/j.compedu.2013.04.013

Drummond, A. & Sweeney, T. (2017). Can an objective measure of technological pedagogical content knowledge (TPACK) supplement existing TPACK measures? *British Journal of Educational Technology*, *48*(4), 928–939. https://doi.org/10.1111/bjet.12473

Dunning, D., Heath, C. & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological science in the public interest : a journal of the American Psychological Society*, *5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Durak, H. Y. & Saritepeci, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education*, *116*, 191–202. https://doi.org/10.1016/j.compedu.2017.09.004

Egloffstein, M. & Ifenthaler, D. (2017). Employee Perspectives on MOOCs for Workplace Learning. *TechTrends*, *61*(1), 65–70. https://doi.org/10.1007/s11528-016-0127-3

Egloffstein, M., Koegler, K. & Ifenthaler, D. (2019). Instructional Quality of Business MOOCs: Indicators and Initial Findings. *Online Learning*, *23*(4). https://doi.org/10.24059/olj.v23i4.2091

Euler, D. (2013). Von der Hochschuldidaktik zur Hochschulentwicklung: neue Herausforderungen für die Gestaltung von Lehre und Studium. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *109*(3), 360–373.

Euler, D. & Sloane, P. F. E. (Hrsg.). (2014). *Zeitschrift für Berufs- und Wirtschaftspädagogik: Beiheft 27*. *Design-based research*. Franz Steiner Verlag.

Faßhauer, U. (2020). Lernortkooperation im Dualen System der Berufsausbildung – implizite Normalität und hoher Entwicklungsbedarf. In R. Arnold, A. Lipsmeier & M. Rohs (Hrsg.), *Handbuch Berufsbildung* (S. 471–484). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-19312-6_37

Finke, S., Kemény, F., Sommer, M., Krnjic, V., Arendasy, M., Slany, W. & Landerl, K. (2022). Unravelling the numerical and spatial underpinnings of computational thinking: a pre-registered replication study. *Computer Science Education*, *32*(3), 313–334. https://doi.org/10.1080/08993408.2022.2079864

Fraillon, J., Ainley, J., Schulz, W., Friedman, T. & Duckworth, D. (2019). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 international report*. International Association for the Evaluation of Educational Achievement (IEA). https://www.iea.nl/publications/study-reports/preparing-life-digital-world

Fraillon, J., Ainley, J., Schulz, W., Friedman, T. & Duckworth, D. (2020). *IEA International Computer and Information Literacy Study 2018: Technical Report*. International Association for the Evaluation of Educational Achievement (IEA). https://www.iea.nl/publications/technical-reports/icils-2018-technical-report

Fütterer, T., Hoch, E., Stürmer, K., Lachner, A., Fischer, C. & Scheiter, K. (2021). Was bewegt Lehrpersonen während der Schulschließungen? – Eine Analyse der Kommunikation im Twitter-Lehrerzimmer über Chancen und Herausforderungen digitalen Unterrichts. *Zeitschrift für Erziehungswissenschaft*, *24*(2), 443–477. https://doi.org/10.1007/s11618-021-01013-8

Ganzeboom, H. B. & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, *25*(3), 201–239. https://doi.org/10.1006/ssre.1996.0010

Gerholz, K.-H. & Sloane, P. F. E. (2011). Lernfelder als universitäres Curriculum? Eine hochschuldidaktische Adaption. *bwp@*(20), 1–23. http://www.bwpat.de/ausgabe20/gerholzsloane-bwpat20.pdf

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology*, *28*(6), 545–579. https://doi.org/10.1037/0022-0167.28.6.545

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education*, *57*(3), 1953–1960. https://doi.org/10.1016/j.compedu.2011.04.010

Guggemos, J., Moser, L. & Seufert, S. (2022). Learners don't know best: Shedding light on the phenomenon of the K-12 MOOC in the context of information literacy. *Computers & Education*, *188*, 104552. https://doi.org/10.1016/j.compedu.2022.104552

Guggemos, J. & Seufert, S. (2018). *Computational Thinking: ein nützliches Konstrukt für die kaufmännische Bildung?!* Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE (September 3 - 5 in Frankfurt a.M.).

Guggemos, J. & Seufert, S. (2021). *Künstliche Intelligenz Literacy – ein wichtiges Konstrukt für die Berufs- und Wirtschaftspädagogik?!* Jahrestagung Sektion Berufs- und Wirtschaftspädagogik der DGfE, September 15- 17 online.

Guggemos, J., Seufert, S. & Román-González, M. (2019). Measuring computational thinking - adapting a performance test and a self-assessment instrument for German-speaking countries.

Proceedings of the 16th International Conference Cognition and Exploratory Learning in the Digital Age (CELDA), 183–191. https://doi.org/10.33965/celda2019_201911L023

Guggemos, J., Seufert, S., Sonderegger, S. & Burkhard, M. (2022). Social Robots in Education: Conceptual Overview and Case Study of Use. In D. Ifenthaler, P. Isaías & D. G. Sampson (Hrsg.), *Orchestration of Learning Environments in the Digital World* (S. 173–195). Springer. https://doi.org/10.1007/978-3-030-90944-4_10

Hair, J. F., Risher, J. J., Sarstedt, M. & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Handke, J. (2018). Digitale Hochschullehre – Vom einfachen Integrationsmodell zur Künstlichen Intelligenz. In U. Dittler & C. Kreidl (Hrsg.), *Hochschule der Zukunft* (S. 249–263). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20403-7_15

Handke, J. (2020). *Humanoide Roboter: Showcase, Partner und Werkzeug* (1. Aufl.). Tectum Wissenschaftsverlag.

Hartig, J. & Frey, A. (2013). Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? [Benefits and limitations of modeling competencies by means of Item Response Theory (IRT)]. *Zeitschrift für Erziehungswissenschaft*, *16*(S1), 47–51. https://doi.org/10.1007/s11618-013-0386-0

Hartig, J., Frey, A., Nold, G. & Klieme, E. (2012). An application of explanatory item response modeling for model-based proficiency scaling. *Educational and Psychological Measurement*, *72*(4), 665–686. https://doi.org/10.1177/0013164411430707

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I. & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, *153*, 103897. https://doi.org/10.1016/j.compedu.2020.103897

Hofmans, J., Wille, B. & Schreurs, B. (2020). Person-centered methods in vocational research. *Journal of Vocational Behavior*, *118*, 103398. https://doi.org/10.1016/j.jvb.2020.103398

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments. 3rd ed*. Psychological Assessment Resources.

Hone, K. S. & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, *98*, 157–168.

Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A. & Yamada, M. (2021). Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners. *Educational*

Technology Research and Development, *69*(4), 2131–2150. https://doi.org/10.1007/s11423-020-09909-8

Ifenthaler, D. & Schweinbenz, V. (2013). *The acceptance of Tablet-PCs in classroom instruction: The teachers' perspectives* (Bd. 29). https://doi.org/10.1016/j.chb.2012.11.004

Ilic, U., Haseski, H. I. & Tugtekin, U. (2018). Publication trends over 10 years of computational thinking research. *Contemporary Educational Technology*, *9*(2), 131–153. https://doi.org/10.30935/cet.414798

ISTE. (2015). *Computational thinking: leadership toolkit*. https://www.iste.org/computational-thinking

Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C. & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research*, *88*(1), 43–86. https://doi.org/10.3102/0034654317740335

J-PAL. (2019). *Will technology transform education for the better?* https://www.povertyactionlab.org/publication/will-technology-transform-education-better

Kay, A. (1971). *The best way to predict the future is to invent it*. https://en.wikiquote.org/wiki/Alan_Kay

Kim, M. C. & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education*, *56*(2), 403–417. https://doi.org/10.1016/j.compedu.2010.08.024

Kirschner, P. A. & Bruyckere, P. de (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, *67*, 135–142. https://doi.org/10.1016/j.tate.2017.06.001

Knezek, G. & Christensen, R. (2016). Extending the will, skill, tool model of technology integration: adding pedagogy as a new model construct. *Journal of Computing in Higher Education*, *28*(3), 307–325. https://doi.org/10.1007/s12528-016-9120-2

Koehler, M. J., Mishra, P. & Cain, W. (2013). What is Technological Pedagogical Content Knowledge (TPACK)? *Journal of Education*, *193*(3), 13–19. https://doi.org/10.1177/002205741319300303

Koller, M. & Stahel, W. A. (2011). Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis*, *55*(8), 2504–2515. https://doi.org/10.1016/j.csda.2011.02.014

Köller, O., Thiel, F., van Ackeren, I., Anders, Y., Becker-Mrotzek, M., Cress, U., Diehl, C., Kleickmann, T., Lütje-Klose, B., Prediger, S., Seeber, S., Ziegler, B., Kuper, H., Stanat, P., Maaz, K. &

Lewalter, D. (2022). *Digitalisierung im Bildungssystem: Handlungsempfehlungen von der Kita bis zur Hochschule. Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK)*. SWK. https://doi.org/10.25656/01:25273

Korkmaz, Ö., Çakir, R. & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, *72*, 558–569. https://doi.org/10.1016/j.chb.2017.01.005

Koutsakas, P., Chorozidis, G., Karamatsouki, A. & Karagiannidis, C. (2020). Research Trends in K–12 MOOCs: A Review of the Published Literature. *The International Review of Research in Open and Distributed Learning*, *21*(3). https://doi.org/10.19173/irrodl.v21i3.4650

Landers, R. N., Brusso, R. C., Cavanaugh, K. J. & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological methods*, *21*(4), 475–492. https://doi.org/10.1037/met0000081

Lehmann, H. & Rossi, P. G. (2019). Social Robots in Educational Contexts: Developing an Application in Enactive Didactics. *Journal of e-Learning and Knowledge Society*, *15*(2), 27–41. https://doi.org/10.20368/1971-8829/1633

Leys, C., Delacre, M., Mora, Y. L., Lakens, D. & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, *32*(1), Artikel 5. https://doi.org/10.5334/irsp.289

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, *77*, 23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001

Long, D. & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In R. Bernhaupt (Hrsg.), *ACM Digital Library, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (S. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Lye, S. Y. & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, *41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Mahmood, K. (2016). Do People Overestimate Their Information Literacy Skills? A Systematic Review of Empirical Evidence on the Dunning-Kruger Effect. *Comminfolit*, *10*(2), 199–213. https://doi.org/10.15760/comminfolit.2016.10.2.24

McKenney, S. & Reeves, T. C. (2014). Educational Design Research. In J. M. Spector, M. D. Merrill, J. Elen & M. J. Bishop (Hrsg.), *Handbook of Research on Educational Communications and Technology* (S. 131–140). Springer New York.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43–59. https://doi.org/10.1007/BF02505024

Meyer, J. P. & Morin, A. J. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior*, *37*(4), 584–612. https://doi.org/10.1002/job.2085

Morin, A. J. S. & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 39–59. https://doi.org/10.1080/10705511.2014.919825

National Academies of Sciences, Engineering, and Medicine. (2018). *How People Learn II*. National Academies Press. https://doi.org/10.17226/24783

Newman, I., Ridenour, C. S., Newman, C. & DeMarco, G. M. P., Jr. (2007). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddlie (Hrsg.), *Handbook of mixed methods in social & behavioral research* (167-188). Sage.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Oberländer, M., Beinicke, A. & Bipp, T. (2020). Digital competencies: A review of the literature and applications in the workplace. *Computers & Education*, *146*, 103752. https://doi.org/10.1016/j.compedu.2019.103752

OECD (2015). Students, computers and learning. Vorab-Onlinepublikation. https://doi.org/10.1787/9789264239555-en

OECD. (2017). *PISA 2015 Technical Report*. OECD Publishing. http://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf

OKGT. (2022). *Berufliche Grundbildung Kauffrau/Kaufmann EFZ Was wird neu auf Lehrbeginn 2023?* https://www.okgt.ch/public/upload/assets/763/20220207_Was_wird_neu_auf_Lehrbeginn_2023_de.pdf

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., Campion, M. A., Mayfield, M. S., Morgeson, F. P., Pearlman, K., Gowing, M. K., Lancaster, A. R., Silver, M. B. & Dye, D. M. (2001). Understanding work using the Occupational Information Network (O*NET): Implications for practice and research. *Personnel Psychology*, *54*(2), 451–492. https://doi.org/10.1111/j.1744-6570.2001.tb00100.x

Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Raffaghelli, J. E., Rodríguez, M. E., Guerrero-Roldán, A.-E. & Bañeres, D. (2022). Applying the UTAUT model to explain the students' acceptance of an early warning system in Higher Education. *Computers & Education*, *182*, 104468. https://doi.org/10.1016/j.compedu.2022.104468

Reeves, T. C. & Lin, L. (2020). The research we have is not the research we need. *Educational Technology Research & Development*, *68*(4), 1991–2001. https://doi.org/10.1007/s11423-020-09811-3

Reich, J. (2015). Education research. Rebooting MOOC research. *Science*, *347*(6217), 34–35. https://doi.org/10.1126/science.1261627

Reich, J. & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, *363*(6423), 130–131. https://doi.org/10.1126/science.aav7958

Repenning, A. (2017). Moving beyond syntax: Lessons from 20 years of blocks programing in AgentSheets. *Journal of Visual Languages and Sentient Systems*, *3*(1), 68–91. https://doi.org/10.18293/VLSS2017-010

Repenning, A., Webb, D. C., Koh, K. H., Nickerson, H., Miller, S. B., Brand, C., Her Many Horses, I., Basawapatna, A., Gluck, F., Grover, R., Gutierrez, K. & Repenning, N. (2015). Scalable game design: A strategy to bring systemic computer science education to schools through game design and simulation creation. *ACM Transactions on Computing Education*, *15*(2), 1–31. https://doi.org/10.1145/2700517

Ringle, C. M. & Sarstedt, M. (2016). Gain more insight from your PLS-SEM results. *Industrial Management & Data Systems*, *116*(9), 1865–1886. https://doi.org/10.1108/IMDS-10-2015-0449

Roll, M. & Ifenthaler, D. (2021). Learning Factories 4.0 in technical vocational schools: can they foster competence development? *Empirical Research in Vocational Education and Training*, *13*(1). https://doi.org/10.1186/s40461-021-00124-0

Román-González, M., Moreno-León, J. & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In S. C. Kong & H. Abelson (Hrsg.), *Computational thinking education* (S. 79–98). Springer. https://doi.org/10.1007/978-981-13-6528-7_6

Román-González, M., Pérez-González, J.-C. & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*, *72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sandoval, W. (2014). Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences*, *23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Sarstedt, M., Hair, J. F. & Ringle, C. M. (2022). “PLS-SEM: indeed a silver bullet” – retrospective observations and recent advances. *Journal of Marketing Theory and Practice*, 1–15. https://doi.org/10.1080/10696679.2022.2056488

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O. & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, *69*(10), 3998–4010. https://doi.org/10.1016/j.jbusres.2016.06.007

Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung. *Zeitschrift für Erziehungswissenschaft*, 1–22. https://doi.org/10.1007/s11618-021-01047-y

Scherer, R., Siddiq, F. & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers’ adoption of digital technology in education. *Computers & Education*, *128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Scherer, R., Siddiq, F. & Tondeur, J. (2020). All the same or different? Revisiting measures of teachers' technology acceptance. *Computers & Education*, *143*, 103656. https://doi.org/10.1016/j.compedu.2019.103656

Scherer, R., Tondeur, J. & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education*, *112*, 1–17. https://doi.org/10.1016/j.compedu.2017.04.012

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J. & Shin, T. S. (2009). Technological Pedagogical Content Knowledge (TPACK). *Journal of Research on Technology in Education*, *42*(2), 123–149. https://doi.org/10.1080/15391523.2009.10782544

Seeber, I., Bittner, E., Briggs, R. O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G. & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174

Sektion BWP der DGfE. (2014). *Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik im Rahmen berufs- und wirtschaftspädagogischer Studiengänge*. http://www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf

Seufert, S., Guggemos, J. & Moser, L. (2019). Digitale Transformation in Hochschulen: auf dem Weg zu offenen Ökosystemen. *Zeitschrift für Hochschulentwicklung*, *14*(2), Artikel 5, 85–107. https://doi.org/10.3217/zfhe-14-02/05

Seufert, S., Guggemos, J. & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, *115*, 106552. https://doi.org/10.1016/j.chb.2020.106552

Seufert, S., Guggemos, J., Tarantini, E. & Schumann, S. (2019). Developing and validating a framework of digital competencies for teachers in the commercial domain Professionelle Kompetenzen von Lehrpersonen im Kontext des digitalen Wandels Entwicklung eines Rahmenkonzepts und Validierung in der kaufmännischen Domäne. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *115*(2), 312–339. https://doi.org/10.25162/zbw-2019-0013

Seufert, S., Slabeva, K. S. & Guggemos, J. (2020). Assessing subjective and objective information literacy at upper secondary schools - an empirical study in four German-speaking countries. *International Journal of Learning Technology*, *15*(1), Artikel 107666, 82. https://doi.org/10.1504/IJLT.2020.107666

Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-Concept: validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.2307/1170010

Shute, V. J., Sun, C. & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, *22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Siddiq, F., Scherer, R. & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): A new construct in 21st century education. *Computers & Education*, *92-93*, 1–14. https://doi.org/10.1016/j.compedu.2015.10.006

Skilling, K. & Stylianides, G. J. (2020). Using vignettes in educational research: a framework for vignette construction. *International Journal of Research & Method in Education*, *43*(5), 541–556. https://doi.org/10.1080/1743727X.2019.1704243

Sloane, P. F. E., Twardy, M. & Buschfeld, D. (2004). *Einführung in die Wirtschaftspädagogik* (2. Aufl.). Eusl-Verl.

Söll, M. (2017). Die Transparenz und Vergleichbarkeit von Studiengängen der Wirtschaftspädagogik im Spiegel hochschulpolitischer Reformbemühungen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *113*(3), 397–427.

Söll, M., Reinisch, H. & Klusmeyer, J. (2014). Publikation und Reputation: Eine Studie zum Rezeptions- und Publikationsverhalten von wissenschaftlich tätigen Berufs- und Wirtschaftspädagoginnen und -pädagogen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *110*(4), 505–528.

Sonderegger, S. (2022). Enhancing learning processes by integrating social robots with learning management systems. In *31st IEEE International Conference on Robot and Human Interactive*

Communication (RO-MAN) (pp. 365–370). IEEE. https://doi.org/10.1109/RO-MAN53752.2022.9900622

Spurk, D., Hirschi, A., Wang, M., Valero, D. & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Stock, M., Slepcevic-Zach, P. & Riebenbauer, E. (2019). Herausforderungen und Chancen einer polyvalenten wissenschaftlichen Berufsvorbildung. In F. Grammlinger, C. Iller, A. Ostendorf, K. Schmid & G. Tafner (Hrsg.), *Herausforderungen und Chancen einer polyvalenten w issenschaftlichen Berufsvorbildung.* wbv. https://doi.org/10.3278/6004660w199

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C. & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning. *Review of Educational Research*, *81*(1), 4–28. https://doi.org/10.3102/0034654310393361

Tang, X., Yin, Y., Lin, Q., Hadad, R. & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, 103798. https://doi.org/10.1016/j.compedu.2019.103798

Tondeur, J., van Braak, J., Sang, G., Voogt, J. M., Fisser, P. & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, *59*(1), 134–144. https://doi.org/10.1016/j.compedu.2011.10.009

Trachsel, V. & Bitterli, C. (2020). Controller-Profile in der Schweiz – Bedeutung der Digitalisierung. In I. Keimer & U. Egle (Hrsg.), *Die Digitalisierung der Controlling-Funktion* (S. 199–210). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29196-9_12

Tramm, T., Kremer, H.-H. & Reinmann, G. (2017). Editorial zur Ausgabe 33: Entwicklungsbezogene (Praxis-)Forschung. *bwp@*(33), 1–5.

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V. & Ninaus, M. (2022). A cognitive definition of computational thinking in primary education. *Computers & Education*, *179*, 104425. https://doi.org/10.1016/j.compedu.2021.104425

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review*, *3*(2), 130–154. https://doi.org/10.1016/j.edurev.2007.12.001

Valtonen, T., López-Pernas, S., Saqr, M., Vartiainen, H., Sointu, E. T. & Tedre, M. (2022). The nature and building blocks of educational technology research. *Computers in Human Behavior*, *128*, 107123. https://doi.org/10.1016/j.chb.2021.107123 van Laar, E., van Deursen, A. J., van Dijk, J. A. & Haan, J. de (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, *72*, 577–588. https://doi.org/10.1016/j.chb.2017.03.010

Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Voogt, J., Fisser, P., Good, J., Mishra, P. & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies*, *20*(4), 715–728. https://doi.org/10.1007/s10639-015-9412-6

Voss, T., Kunina-Habenicht, O., Hoehne, V. & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde: Empirische Zugänge und Befunde. *Zeitschrift fur Erziehungswissenschaft*, *18*(2), 187–223. https://doi.org/10.1007/s11618-015-0626-6

Vuorikari, R., Kluzer, S. & Punie, Y. (2022). *DigComp 2.2 - The Digital Competence Framework for Citizens*. Publications Office of the European Union. https://doi.org/10.2760/115376

Weber, S., Trost, S., Wiethe-Körprich, M., Weiss, C. & Achtenhagen, F. (2014). Intrapreneur: An Entrepreneur within a Company: An Approach on Modeling and Measuring Intrapreneurship Competence. In S. Weber, F. K. Oser, F. Achtenhagen, M. Fretschner & S. Trost (Hrsg.), *Professional and Vet Learning. Becoming an Entrepreneur* (S. 279–302). SensePublishers.

Weintrop, D., Wise Rutstein, D., Bienkowski, M. & McGee, S. (2021). Assessing computational thinking: an overview of the field. *Computer Science Education*, *31*(2), 113–116. https://doi.org/10.1080/08993408.2021.1918380

Wenner, T. (2018). Entwicklung eines Instruments zur Erfassung der Wechselwirkung von Lernortkooperation und Ausbildungsqualität. *Journal of Technical Education (JOTED)*, *6*(1), 223–237. https://doi.org/10.48513/joted.v6i1.123

Whetten, D. A. (1989). What Constitutes a Theoretical Contribution? *Academy of Management Review*, *14*(4), 490–495. https://doi.org/10.5465/amr.1989.4308371

Wigfield, A., Muenks, K. & Eccles, J. S. (2021). Achievement motivation: What we know and where we are going. *Annual Review of Developmental Psychology*, *3*(1), 87–111. https://doi.org/10.1146/annurev-devpsych-050720-103500

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions. Series A, Mathematical, Physical, And Engineering Sciences*, *366*(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118

Wong, G. K. W., Ma, X., Dillenbourg, P. & Huan, J. (2020). Broadening artificial intelligence education in K-12. *ACM Inroads*, *11*(1), 20–29. https://doi.org/10.1145/3381884

Woo, H., LeTendre, G. K., Pham-Shouse, T. & Xiong, Y. (2021). The use of social robots in classrooms: A review of field-based studies. *Educational Research Review*, *33*, 100388. https://doi.org/10.1016/j.edurev.2021.100388

XING AG. (2022). *Geschäftsbericht 2016*. https://corporate.xing.com/de/investor-relations/publikationen/

Yadav, A., Good, J., Voogt, J. & Fisser, P. (2018). Computational thinking as an emerging competence domain. In M. Mulder (Hrsg.), *Technical and Vocational Education and Training. Competence-based Vocational and Professional Education: Bridging the Worlds of Work and Education* (S. 1051–1067). Springer International Publishing. https://doi.org/10.1007/978-3-319-41713-4_49

Yadav, A., Hong, H. & Stephenson, C. (2016). Computational thinking for all: Pedagogical approaches to embedding 21st century problem solving in K-12 classrooms. *TechTrends*, *60*(6), 565–568. https://doi.org/10.1007/s11528-016-0087-7

Zawacki-Richter, O., Marín, V. I., Bond, M. & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0171-0

Zhao, W. & Shute, V. J. (2019). Can playing a video game foster computational thinking skills? *Computers & Education*, *141*, 103633. https://doi.org/10.1016/j.compedu.2019.103633

Zhu, M., Sari, A. R. & Lee, M. M. (2020). A comprehensive systematic review of MOOC research: Research techniques, topics, and trends from 2009 to 2019. *Educational Technology Research and Development*, *68*(4), 1685–1710. https://doi.org/10.1007/s11423-020-09798-x

Studie 1:

Guggemos, J., Seufert, S. & Román-González, M. (2023). Computational Thinking Assessment – Towards More Vivid Interpretations. *Technology, Knowledge and Learning, 28*, 539–568.

Technology, Knowledge and Learning (2023) 28:539–568
https://doi.org/10.1007/s10758-021-09587-2

ORIGINAL RESEARCH

Computational Thinking Assessment – Towards More Vivid Interpretations

Josef Guggemos[1] · Sabine Seufert[1] · Marcos Román-González[2]

Accepted: 10 December 2021 / Published online: 27 January 2022

Abstract

Computational thinking (CT) is an important 21st-century skill. This paper aims at more useful CT assessment. Available evaluation instruments are reviewed; two generally accepted CT evaluation tools are selected for a comprehensive CT assessment: the CTt, a performance test, and the CTS, a self-assessment instrument. The sample comprises 202 high school students from German-speaking Switzerland. Concerning the CTt, Rasch-scalability is demonstrated. Utilizing the approach of the PISA studies, proficiency levels are formed that comprise tasks with specific characteristics that students are systematically able to master. This could help teachers to offer individual support to their students. In terms of the CTS, the original version is refined using confirmatory factor and measurement-invariance analysis. A latent profile analysis yielded four profiles, two of which are of particular interest. One profile comprises students with, on the one hand, moderate to high creative thinking ability, cooperativity, and critical thinking skills and, on the other hand, low algorithmic thinking ability. The second remarkable profile consists of students with particularly low cooperativity. Based on these strength and weakness profiles, teachers could offer support tailored to student needs.

Keywords Computational thinking · Performance test · Item response theory · Latent profile analysis · Person-centered assessment · Proficiency level model

Abbreviations

aBIC	Adjusted Bayesian information criterion
AIC	Akaike's information criterion
BIC	Bayesian information criterion
BLRT	Bootstrap likelihood ratio test
CAIC	Consistent Akaike's information criterion
CFI	Comparative fit index
CI	Confidence interval

✉ Josef Guggemos
josef.guggemos@unisg.ch

[1] TechnologiesSchool of Management, Institute forEducational Management and Technologies, University of St.Gallen, St. Jakob-Strasse 21, 9000 St. Gallen, Switzerland

[2] Universidad Nacional de Educación a Distancia (UNED), Faculty of Education, C/ Juan del Rosal 14, C.P. 28040 Madrid, Spain

CT	Computational thinking
CTS	Computational Thinking Scale
CTt	Computational Thinking Test
DIF	Differential item functioning
EAP/PV	Expected a posteriori / plausible values
ICILS	International Computer and Information Literacy Study
LPA	Latent profile analysis
LRT	Andersen's likelihood ratio test
MANOVA	Multivariate analysis of variance
MLR	Maximum likelihood with robust standard errors
PISA	Programme for International Student Assessment
R^2	Coefficient of determination
RMSEA	Root mean square error of approximation
RQ	Research question
SRMR	Standardized root mean square residual
TLI	Tucker-Lewis Index
WLE	Weighted likelihood estimate
WLSMV	Diagonally weighted least squares
wMNSQ	Weighted mean square
YB-χ^2	Yuan-Bentler corrected χ^2
α	Cronbach's alpha
ω	Revell's omega total

1 Introduction

Tremendous technological changes are shaping our society and ways of working (Harteis et al., 2020; Ifenthaler et al., 2021; Kirschner & Stoyanov, 2020). In this context, computational thinking (CT) is regarded as a key 21st-century skill (Voogt et al., 2015; Wing, 2006; Yadav et al., 2016). The significance of CT in the twenty-first century may be evident (Barr et al., 2011; Buitrago Flórez et al., 2017; Wing, 2008): computing and computer technology pervade every field of study and workplace. CT aims at enabling humans to use these resources for solving problems.

In her seminal paper, Wing conceptualizes CT as "solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science" (Wing, 2006, p. 33). As the theoretical basis for CT, the framework of Brennan and Resnick (2012) is often utilized. It comprises three dimensions (Brennan & Resnick, 2012, pp. 3–11). *Computational concepts* are common in many programming languages, but not restricted to them: sequences, loops, events, parallelism, conditionals, operators, and data. *Computational practices* are the processes in which students engage: imaging and building, testing and debugging, reusing and remixing, abstracting, and modularizing. *Computational perspectives* are shifts in points of view, relationships to others, and the digital world around: expressing, connecting, and questioning.

In CT research, the relationship between CT and programming is often thematized. Israel et al. (2015) regard the use of computers to model ideas and programming as an integral part of CT. Buitrago Flórez et al. (2017), as well as Lye and Koh (2014), argue that by means of programming, several core facets of CT can be addressed. Shute et al. (2017) concluded that there is a close relationship between CT and programming skills due to similar underlying cognitive processes. Hsu et al. (2018), based on their review of the literature, reported that programming is widely used to teach CT. Grover et al. (2016) maintained that programming has a positive influence on the experience on CT. Scherer et al. (2019), based on a meta-analysis, concluded that CT can be taught through programming. However, using professional programming languages like Java can be extremely difficult for students due to complex syntax, and it may be preferable to use visual programming languages (Lye & Koh, 2014; Repenning, 2017). Scratch, developed by the Massachusetts Institute of Technology Media Lab (https://scratch.mit.edu), is such a visual programming language that is heavily used as an instructional tool (Hsu et al., 2018).

The importance of CT assessment is regularly stressed (Grover & Pea, 2013; Ilic et al., 2018; Shute et al., 2017; Tang et al., 2020; Weintrop et al., 2021). However, it has to be highlighted that assessment is not an end in itself, but it should contribute to promoting student learning (Pellegrino et al., 2016). When assessing complex skills such as CT, the structure as well as the levels of the construct have to be considered (Seufert et al., 2021). For instance, the DigComp 2.1 framework that addresses digital competencies comprises five dimensions (structure), e.g., information literacy, as well as eight proficiency levels ranging from foundation to highly specialized (Carretero et al., 2017). In terms of CT, research about proficiency levels is in its infancy. The importance of modeling proficiency levels for a better understanding of this construct, however, is stressed by the 2018 International Computer and Information Literacy Study (ICILS) (Fraillon et al., 2019). Without proficiency levels, test results are difficult to comprehend and to communicate. Findings such as 'the student belongs to the top 10% of all test takers' or 'the student answered 67% of the questions correctly' are not very helpful for the purpose of fostering student learning. A proficiency level model could allow for a more vivid interpretation of the test results by relying on items with clearly specified characteristics that students are systematically able to master (AERA et al., 2014). This also makes it possible to set operationalized learning goals. For example, the goal could be that all students in a class are systematically able to use functions. Students who do not reach the corresponding proficiency level could receive specific guidance and support.

Due to the multifaceted nature of CT, it is unlikely that a single instrument is sufficient to comprehensively capture CT (Polat et al., 2021; Román-González et al., 2019). Rather, a system of various assessments may be necessary. Using multidimensional approaches could also reveal different CT profiles. For instance, some students may perform well in the realm of computational concepts, but at the same time poorly in the area of computational perspectives. By solely focusing on a specific facet of CT, students with specific strength and weakness profiles may not be identified. Knowledge about CT profiles, i.e., different types of computational thinkers, however, could be utilized for personalized guidance and support (Hofmans et al., 2020). Although such person-centered assessments are widespread in other research fields (e.g., Lohr et al., 2021; Meyer & Morin, 2016; Scherer et al., 2021; Tondeur et al., 2019), they are not common in CT research. The benefits of such techniques, however, are acknowledged in CT research (Román-González et al., 2019).

2 Theoretical Perspectives

2.1 CT Assessment Instruments

Several authors have reviewed CT assessment instruments (Kong, 2019; Román-González et al., 2019; Shute et al., 2017; Tang et al., 2020). Román-González et al. (2019) developed a useful classification of assessment instruments (Israel-Fishelson & Hershkovitz, 2022). Following this classification, diagnostic tools that aim at capturing students' CT proficiency could be the most suitable basis for forming proficiency levels. Diagnostic tools are performance tests and do not require specific prior knowledge, e.g., a specific programming language. Hence, they can be used to evaluate learning gains by comparing pre- and post-test results, i.e., if students have reached a higher proficiency level after instruction.

A drawback of many diagnostic tools is that they are not freely available (ICILS 2018: Fraillon et al., 2019; Fairy Assessment: Werner et al., 2012; Basic Programming Abilities: Mühling et al., 2015). Furthermore, several tools focus on the pre-secondary level (Chen et al., 2017; Kong & Wang, 2021; Relkin et al., 2021; Seiter & Foreman, 2013); therefore, the covered proficiency spectrum is limited. From the freely available diagnostic tools, the Computational Thinking Test (CTt) may be especially suitable for the purpose of forming proficiency levels. The CTt (Román-González, 2015) is a performance test for secondary students using the framework of Brennan and Resnick (2012) as a theoretical background. It defines CT as "the ability to formulate and solve problems by relying on the fundamental concepts of computing, and using logic-syntax of programming languages: basic sequences, loops, iteration, conditionals, functions and variables" (Román-González et al., 2017, p. 681). Sample items can be found in Figs. 2 and 3. Due to the use of a visual programming language, the CTt covers a broad range of instructional settings. To perform the CTt no knowledge in a specific programming language, e.g., Java, is necessary, which makes it a very flexible instrument. The CTt comprises 28 selected response items and can be taken online; the target group should be able to process the test in less than 45 min. The CTt is claimed to be unidimensional although different cognitive operations are involved when performing the items. This is based on the notion of Fischer (1973) that the items of a unidimensional construct may be linearly decomposed into cognitive operations; this may also be the case for CT (Mühling et al., 2015). According to the definition of the CTt, cognitive operations could be *sequences*, *loops*, *conditionals*, *functions*, and *variables*. These correspond with the computational concepts dimension of the Brennan and Resnick (2012) framework.

Román-González et al., (2017) validated the CTt using a sample of 1,251 Spanish secondary students (5th to 10th grade) and classical test theory. The reliability of the test is sufficiently high (Cronbach's alpha = 0.79). Chan et al. (2020) provided evidence for Rasch scalability of the CTt based on a sample of 153 upper-secondary students from Singapore. The CTt is increasingly used in research projects for assessing CT learning (e.g., Guggemos, 2021; Brackmann et al., 2017; Hooshyar et al., 2021; Rose et al., 2019; Zhao & Shute, 2019). In light of this, relying on the CTt may be in line with the call for using standardized instruments to ensure comparability across studies (Shute et al., 2017).

Despite its suitability for assessing computational concepts and, to some degree, computational practices, the CTt also has disadvantages: it neglects computational perspectives (Román-González et al., 2017). For capturing such perspectives, perception-attitude scales may be suitable (Román-González et al., 2019). They capture self-efficacy beliefs by means of self-assessment. In general, complementing a performance test with a self-assessment instrument may be beneficial to obtain a comprehensive picture of a construct (Rosman et al., 2015). Again, we aim at using standardized instruments and those with a specific focus on CT. This excludes computer attitude scales, e.g., Denner et al. (2014), Ericson and McKlin (2012), and Yadav et al. (2014). For the same reason, we do not consider generic self-efficacy or motivation scales. A viable option to capture computational perspectives may be the Computational Thinking Scales (CTS) (Polat et al., 2021). They were developed by Korkmaz et al. (2017) and is a standardized self-assessment instrument for capturing CT (Durak & Saritepeci, 2018; Israel-Fishelson & Hershkovitz, 2022; Shute et al., 2017). The authors utilize the International Society for Technology in Education (ISTE, 2015) framework of CT, namely, the five dimensions of *creativity*, *algorithmic thinking*, *cooperativity*, *critical thinking*, and *problem solving*. Descriptions of these dimensions can be found in Table 1.

The CTS consists of 29 self-assessment questions. It has been validated by means of confirmatory factor analysis using a sample of 580 Turkish undergraduate students. Fit-values are decent (Korkmaz et al., 2017, p. 565): CFI = 0.95, RMSEA = 0.06.

Overall, the CTt and the CTS may be complementary assessment tools that can provide a comprehensive picture of students' CT ability (Polat et al., 2021). The CTt, as a unidimensional performance test, seems to be suitable to form proficiency levels, especially because the cognitive operations are documented. The CTS, as a multidimensional self-assessment instrument, can be used to identify CT profiles.

2.2 The Present Study

The current study aims at reaching a better understanding of CT as a construct. To this end, we contribute to more useful interpretations of CT assessment findings. Concerning performance tests, the importance of proficiency levels is stressed (Fraillon et al., 2019; OECD, 2017). However, the 2018 ICILS refrained from developing a proficiency level model due to the small number of CT test items in the study (Fraillon et al., 2020); only proficiency regions were described. When forming proficiency levels, referring to the cognitive operations involved when performing the items could be advantageous from a construct validity point of view (Embretson & Daniel, 2008). For instance, it would not be meaningful to form CT proficiency levels based on the text complexity of the items. In this regard, we can take advantage of the fact that for the CTt, the cognitive operations necessary for performing the items are documented. As the cognitive operations correspond with the computational practices of the Brennan and Resnick (2012) framework, this could be a sound theoretical basis. The paper at hand demonstrates that the cognitive operations can predict the difficulty of CT test items and can be utilized to form meaningful proficiency levels (Hartig et al., 2012). Our first research question is:

RQ1 What CT proficiency levels can be identified based on the CTt?

As already highlighted, the purpose of an assessment is to facilitate student learning. In the case of CT as a multifaceted construct, it is unlikely to observe only one CT profile. Rather, it can be expected to identify different types of computational thinkers. For teachers, knowledge about CT profiles could be helpful in reducing complexity. Students in the same profile could benefit from similar treatment (Hofmans et al., 2020); teachers may design their instructional measures around the identified profiles. Despite their usefulness for a better understanding of CT, to our knowledge, such person-centered methods of assessment have not yet been used in CT research. The CTS may be suitable for identifying latent profiles as it comprises five dimensions. Our second research question is:

RQ2 What CT profiles can be identified based on the CTS?

The CTt and CTS are regarded as complementary instruments that can offer a comprehensive picture of student CT (Polat et al., 2021; Román-González et al., 2019). Against this backdrop, we consider a third research question:

RQ3 How are the CTt and CTS results related to each other?

Since, recently, Polat et al. (2021) addressed the same research question, we will compare their findings with those of the present study.

3 Method

3.1 Sample and Adaption of Instruments

Our sample comprises 202 upper-secondary students from German-speaking Switzerland. They all attended the 11th (second last) grade at a *Kantonsschule* (high school), which is the most demanding school type in Switzerland. Data were collected at the beginning of the school year 2018/19. The CTt, CTS, and context questions were administered using Unipark. Teachers supervised the students and ensured an adequate test environment, e.g., preventing copying from their neighbor. The intended test time was 90 min. Ninety-five percent of the students were able to finish the test within this time; teachers allowed every student to complete the work. On average, the students were 17.23 years old (SD = 0.85 years) and 56% were female. They experienced, on average, 2.89 h (SD = 1.20 h) of computer science instruction in the past; students reported tigerjyton (https://www.tigerjython.ch/engl/) as the most generally used learning environment. Tigerjyton addresses important computational concepts such as sequencing, conditionals, functions, and loops. Overall, 77% of the students claim to be able to program, e.g., in Java or Python. To evaluate test motivation, we can draw information from two items within the context questionnaire: 'When performing the tasks, I disengaged' and 'My mind was elsewhere when I was performing the tasks' (Prenzel et al., 1998). Cronbach's alpha equals 0.78. On average, the students disagreed concerning a lack in test motivation: M = 3.36 and SD = 1.66, based on a seven-point scale of rating. This is consistent with an absence of missing data; the omission of items can act as a proxy for a lack of test motivation (Ulitzsch et al., 2020). We also checked for multivariate outliers using Mahalanobis distances (Leys et al., 2019); no student is classified as an outlier at the 1% (and 5%) significance level.

Since the CTt was designed for 5th to 10th grade students, we replaced the five easiest items by five equivalent but more difficult ones. Equivalent means comparable in the environmental interface (canvas vs. maze), answer style, and required task. To this end, we drew on the initial pool of CTt items, namely forty, as well as where experts gauged an item difficulty (Román-González, 2015). The replacement of very easy items may be advantageous as they do not have an evidentiary value (Köhler & Hartig, 2017). A pre-test indicated that almost all the students within the target group (11th grade) would master the easiest five items. Figure 4 depicts an item that was integrated within our German CTt version in comparison to the original version. The item numbering in this paper always refers to our German version; the numbering allocation of the original CTt version can be found in Table 2. To adapt the CTS from English to German we applied a back-translation approach (Maneesriwongul & Dixon, 2004).[1]

3.2 RQ1: Forming Proficiency Levels

3.2.1 Psychometric Test Validation

Before forming proficiency levels, a psychometric test validation is necessary. If the CTt was Rasch scalable, this would imply specific objectivity: students (proficiency) and items (difficulty) can be located on a common Logit scale. This allows for a criterion-referenced test interpretation (Hartig & Frey, 2013). If the proficiency of a person equals the difficulty of an item (same location on the Logit scale), the expected probability of a correct response will be 50%. The proficiency of students can be described by referring to items that they are expected to master with a specified probability.[2] Specific objectivity would be violated, for instance, if some items were more difficult for males than for females.

For assessing Rasch scalability, we draw on the framework of Bühner (2011, p. 547). First, we carry out Andersen's likelihood ratio test (LRT) (Andersen, 1973) using the R-package 'eRm 1.0–1' (Mair & Hatzinger, 2007). A significant LRT would indicate that the items work differently in specific subgroups, i.e., different parameter estimations for the difficulty of the items are obtained. In order to perform the LRT, the students in the sample have to be split up into subgroups. We use the median of the CTt raw score, gender (male vs. female), age (above and below average), computer literacy (above and below average), and ability to program (yes vs. no) as split criteria (Chan et al., 2020; Guggemos et al., 2019). Computer literacy is captured by the dimension *practical computer knowledge* of the INCOBI-R (Richter et al., 2010), and ability to program via student self-reporting. In case of a significant LRT, the next step is to check which items work differently in the subgroups. For instance, students who are able to program may have a systematic advantage in answering specific items (DIF-effect). Based on DIF-analyses ('TAM 3.5–19' package in R; Robitzsch et al., 2020), we may exclude items that systematically advantage or disadvantage specific subgroups. Test fairness (absence of DIF) is an important characteristic of an assessment instrument (AERA et al., 2014). In our case, it is of specific importance as it is the prerequisite for locating all students on one logit scale. In line with Penfield and Algina (2006, pp. 307–308), a DIF of less than 0.43 Logit may be negligible, between 0.43 and 0.64 Logit moderate, and above 0.64 Logit large.

The LRT and DIF-analyses rely on pre-specified split criteria. However, there may also be latent subpopulations of individuals for which the CTt works differently or who show deviant response behavior. Such latent subpopulations can be identified with a mixed Rasch analysis ('mixRasch' 1.1 package in R; Willse, 2011). An example of a latent subgroup could be students who are guessing in order to solve the selected response CTt items. If the mixed Rasch analysis reveals a one-class solution, this would be evidence for the overall fairness of the CTt. To identify the optimal number of latent classes, information criteria are used. They consider goodness of fit and

[1] The German version of the CTt and CTS are available from the authors upon request.

[2] The Rasch model predicts as the probability for a correct answer: exp(proficiency–difficulty) / (1 + exp(proficiency–difficulty)). For instance, a person with a proficiency of 1.5 Logit is expected to solve an item with a difficulty of 1.0 Logit with a probability of 62.2%.

penalize model complexity. We compare models with 1 to 6 latent classes and select the one with the lowest Akaike's information criterion (AIC), as recommended by Bühner (2011, p. 547).

Besides person homogeneity, unidimensionality is necessary to justify the allocation of students and all items on a common Logit scale. The CTt is designed to be unidimensional. Hence, we do not have any assumptions about meaningful factors, other than CT driving students' response behavior. To check for unidimensionality, we relied on confirmatory factor analysis ('lavaan 0.6–7' package in R; Rosseel, 2012). Since the data are ordinal (correct/incorrect), we applied a WLSMV-estimator (Li, 2016). A chi-square test acts as a global fit test. Furthermore, we rely on CFI, TLI, RMSEA, and SRMR as fit measures. Cut-off values for a decent fit may be: CFI and TLI > 0.95, RMSEA < 0.08, and SRMR < 0.11 (Bühner, 2011, pp. 425–427). Poor fit measures would indicate omitted factors that drive the response behavior. For example, if different answer styles were used (see Figs. 3 and 4), this could explain (besides CT) the answer behavior.

The linear logistic test model, as an extension of the Rasch model (Fischer, 1973), allows us to assess whether the cognitive operations involved in the CTt, e.g., sequencing, can explain a substantial proportion of item difficulty. This may be the prerequisite for forming proficiency levels based on the cognitive operations (AERA et al., 2014). A proportion of explained variance (R^2) of 26% might be the minimum that justifies the use of the cognitive operations for forming proficiency levels (Hartig et al., 2012). We utilize the 'eRm 1.0–1' package in R to estimate the linear logistic test model.

After having checked Rasch scalability, we examine if the items meet the cut-off values applied in the PISA studies (OECD, 2017, pp. 131–134; OECD, 2015, pp. 148–151). The deviance from the item discrimination implied by the Rasch model is evaluated by means of the weighted mean square error (wMNSQ = Infit). Discrimination, along these lines, means to separate students in terms of their CT proficiency. For example, if all students were able to master a certain item, this item would have no discriminatory power. The wMNSQ should lie between 0.8 and 1.2; however, wMNSQ values up to 1.33 might be acceptable (Wilson, 2005, p. 129). Items above the upper limit have a too low discrimination, whereas items below the lower limit have a too high discrimination. The point-biserial correlation is a measure for item discrimination in classical test theory and should be above 0.30. The percentage of correct answers should fall between 20 and 90%. Not more than 10% of missing data should be present.

3.2.2 Building Proficiency Levels

For a criterion-referenced interpretation of the CTt results, we form proficiency levels utilizing the approach in the PISA studies (OECD, 2017, pp. 276–287). To this end, we split up the continuum of CT. We choose a width of 1.0 logits for the proficiency levels and a response probability of 62%. This means that students at the bottom of a proficiency level are expected to solve items at the bottom of that level with a probability of 62%, and at the top of the level with a probability of 38%. We opted for a width of 1.0 logits for the proficiency levels, instead of 0.8 as used in the PISA studies, because the manifestation of the cognitive operations indicates a width of one logit, which is permitted (OECD, 2017, p. 281). For every proficiency level, we provide an anchor item. These items are located about 0.5 Logit below the start of the respective level on the Wright map, corresponding with a response probability of about 62%.

3.3 RQ2: Identifying CT Profiles

3.3.1 Psychometric Test Validation

Before identifying latent profiles, we have to evaluate the psychometric properties of the CTS. Since we measured the items on a seven-point scale of rating, ranging from 'not true at all' to 'entirely true', we utilize confirmatory factor analysis with an MLR-estimator (Robitzsch, 2020). To assess the overall model fit, we use CFI, TLI, RMSEA, and SRMR. Moreover, we check for convergent and discriminant validity relying on the average variances extracted and the heterotrait–monotrait ratio. An average variance extracted greater than 0.5, and a heterotrait–monotrait ratio smaller than 0.85, indicate sufficient convergent and discriminant validity (Hair et al., 2019). An average variance extracted above 0.5 implies that more than 50% of the item variance can be explained by the corresponding factor and less than 50% is error variance. A heterotrait–monotrait ratio below 0.85 indicates that the used items capture empirically distinguishable constructs. Revell's omega total (ω) acts as a measure for internal consistency reliability because it is superior to Cronbach's alpha (α) (McNeish, 2018). Since α is widely used, however, we report it along with ω.

Analogous to the DIF-analyses for the CTt, we have to assess the measurement invariance of the CTS, i.e., a similar meaning of the constructs among subgroups, e.g., among males and females. To control for measurement invariance we apply the approach of van de Schoot et al. (2012). Since our aim is to build latent classes based on manifest means, we have to demonstrate full uniqueness measurement invariance. We use a likelihood ratio test to compare the unrestricted model where all parameters are freely estimated with a model where loadings, intercepts, and error variances across groups are restricted to be equal. In line with the DIF-analyses for the CTt, we form groups based on gender, age, computer literacy, and ability to program.

3.3.2 Identifying Latent Profiles

We use the 'tidyLPA 1.0.8' R-package in combination with MPlus 8 to identify student CT profiles by means of a latent profile analysis (LPA) (Hallquist & Wiley, 2018; Rosenberg et al., 2018). We apply an MLR-estimator (Scherer et al., 2021); missing data are not present. In light of our rather small sample size, we have to restrict variances to be equal across profiles and the covariance among the variables to be zero in order to achieve convergence (Meyer & Morin, 2016). The critical step in the LPA is to identify an appropriate number of profiles. This decision might be based on information criteria and likelihood ratio tests, as well as on conceptual deliberations (Scherer et al., 2021). Against this backdrop, we first assessed different class solutions. Following Morin and Marsh (2015) and Hofmans et al. (2020), we report the information criteria AIC, CAIC, BIC, and aBIC, as well as the bootstrap likelihood ratio test (BLRT). Since these criteria may point to a different number of optimal profiles, we also utilize the approach of Akogul and Erisoglu (2017) where information criteria are weighted to determine the optimal number of latent profiles (from an empirical point of view). The number of constructs in the LPA could be a reasonable maximum for the number of latent classes (Tondeur et al., 2019), i.e., if four constructs are considered in the LPA, four profiles could be the maximum. The identified solution should have a sufficiently high precision of classification, indicated by an entropy greater than 0.7. However, the entropy should not be used as a model selection criterion (Sarstedt et al., 2011). To demonstrate the robustness of the findings we conduct a replication of the LPA with 100 bootstrap samples of 150 students from our overall sample of 202 students (Vanslambrouck et al., 2019). Besides this, the profiles should be substantially different from each other, which can be checked by means of a MANOVA (Tondeur et al., 2019). Afterwards, we evaluate if this approach yields a meaningful solution. The profiles should be of reasonable size and show substantial shape differences, i.e., specific strength and weakness profiles that not only differ in levels but also in their pattern (Morin & Marsh, 2015).

4 Results

4.1 RQ1: CT Proficiency Levels

4.1.1 Psychometric Validity of the CTt (German Version)

Of the 28 items, the students in the sample answered on average 18.45 items correctly (SD = 5.71, median = 19, min = 6, max = 28). Concerning Rasch scalability, the LRT yielded mixed results. We did not find significant DIF-effects in terms of gender ($\chi^2 = 36$, df = 27, $p = 0.11$), age ($\chi^2 = 16$, df = 26, $p = 0.94$), or computer literacy ($\chi^2 = 30$, df = 26, $p = 0.26$). However, utilizing the median of the CTt score and ability to program as a split criterion yielded significant DIF: $\chi^2 = 77$, df = 27, $p < 0.01$ and $\chi^2 = 48$, df = 25, $p < 0.01$, respectively. Four items caused this overall DIF-effect. Item 1 was far too easy for the students in our sample (-4.80 Logits) and therefore has no discriminatory power. Item 10 may have caused problems due to a different response format. The provided answer 'Options A and C are correct' might have confused students: many high-achieving students selected Option A. Items 1 and 10 can be found in Appendix 1. For items 11 and 20, we could not find a reason on the content level. Moreover, the DIF-effects are only light to moderate: Logit = 0.58 and 0.51, respectively. Against this background and considering content validity in terms of alignment with the framework of Brennan and Resnick (2012), we decided to exclude items 1 and 10 from the test and retain items 11 and 20. All further analysis was carried out without items 1 and 10, i.e., with 26 items.

The mixed Rasch analysis revealed a one-class solution; the AIC is lower in comparison to any multiclass solution (e.g., AIC for one class = 4944, AIC for two classes = 4982, and AIC for three classes = 5024).

The assumption of unidimensionality (item homogeneity) of the CTt is justified. The CFA with the 26 items loading on a single factor showed a decent fit: $\chi^2(199) = 341$ ($p = 0.049$), CFI = 0.964, TLI = 0.961, RMSEA = 0.026 (90% CI [0.000, 0.039]), SRMR = 0.063.

The linear logistic test model indicated the following cognitive operations as predictors of CTt item difficulty: sequencing: 1.82 Logit, 95% CI [1.46, 2.18], conditionals: 0.24 Logit, 95% CI [0.09, 0.39], functions: 0.77 Logit, 95% CI [0.54, 0.99], and variables: 2.02 Logit, 95% CI [1.77, 2.02]. Overall, these characteristics can explain 62.8% of the difficulty variance of the 26 items, which is well above the minimum acceptable value of 26%.

Concerning the cut-off values from the PISA studies, in general, all items show good values. The wMNSQ lies between 0.89 and 1.15 with the exception of item 18. This item has a wMNSQ of 1.22, which is slightly above the cut-off value of 1.2, but below 1.33. All point-biseral correlations are higher than 0.30. The percentage of correct answers for all

Logit	Student	CTt item	Level
3.5			
	XXXXXX		
3.0			
2.5			
	XXXXXXXXXX		
2.0			III
1.5	XXXXXXXXXXXXXXX		
	XXXXXXXXXXXXXXXX	18	
1.0			
	XXXXXXXXXXXXX		
0.5	XXXXXXXXXX	27	
	XXXXXXXXXXX		
0.0	XXXXXXXX		
	XXXXXXXXXXXXXXXXXXXXXXXXXXX	7 17 **28**	II
-0.5	XX	4 5 26	
	XXXXXXXXX	11 20 24	
-1.0	XXXXXXXXXXXXXXXXXXXXXXXXX	12 14 21 22	
	XXXXXXXXXXXXX	**8**	I
-1.5	XXXXXXXXXX	13 23 25	
	XXXXXXXXXXXX	9 16	
-2.0	XXXXXX	19	
	XXXX	**6** 15	
-2.5	XX	2	below I
	XX		
-3.0	X	3	
-3.5			

Fig. 1 Wright Map of CTt and proficiency levels

items lies between 90 and 25%. Every student fully processed the items; missing values are not present. Table 2 summarizes the item characteristics.

As the Wright map (see Fig. 1) indicates, the items are slightly too easy for the students in our sample. Nevertheless, the EAP/PV-reliability equals 0.85, WLE-reliability 0.81, which is sufficiently high for research purposes. However, if more difficult items were used, we could expect an even higher reliability.

4.1.2 Proficiency Levels

The proficiency levels are illustrated in Fig. 1. All anchor items, in bold, have negligible DIF-effects (<0.41 Logit).

Level I ($-1.75 \leq \text{Logit} < -0.75$) and below: Level I is characterized by tasks using a flow structure without further elements, like conditionals or variables. Item 6 is an anchor item for this level (see Fig. 2). Of the students in the sample, 7.4% do not achieve level I; hence, they are systematically unable to perform sequencing tasks.

For younger students, it could be necessary to insert a level below level I, which contains simple flow structures. Students in our sample, however, solved tasks like the

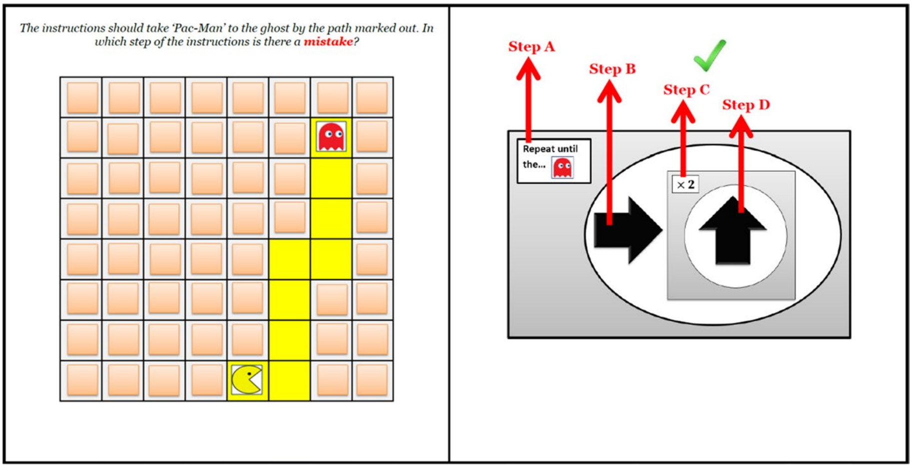


Fig. 2 Item 6, containing sequencing

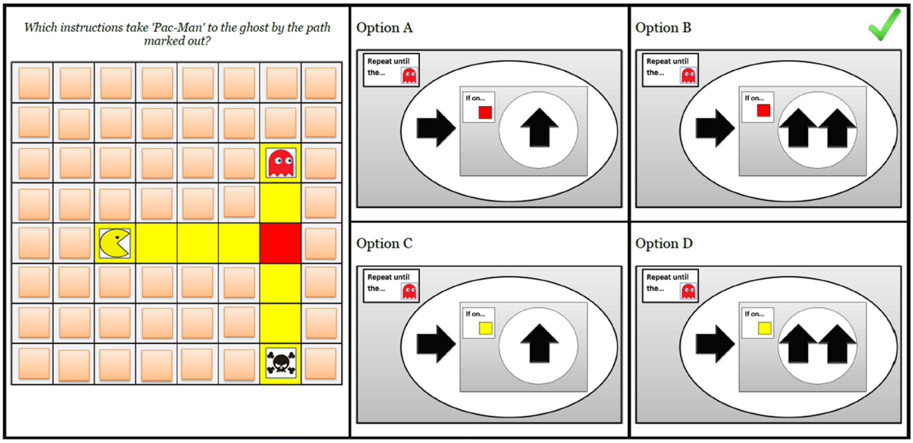


Fig. 3 Item 8, containing sequencing and conditionals

Table 1 CT dimensions (Korkmaz et al., 2017)

CT dimension	Description
Creativity	Coming up with new products; carrying out tasks in new ways, developing new ideas, finding new solutions, taking new viewpoints
Algorithmic thinking	Formalizing the solution of a problem, using a step-by-step approach to come to a solution
Cooperativity	Collaboration with others for a specific purpose
Critical thinking	Testing the reliability of information, avoiding cognitive errors, asking questions
Problem solving	Overcoming obstacles to achieve a specific goal

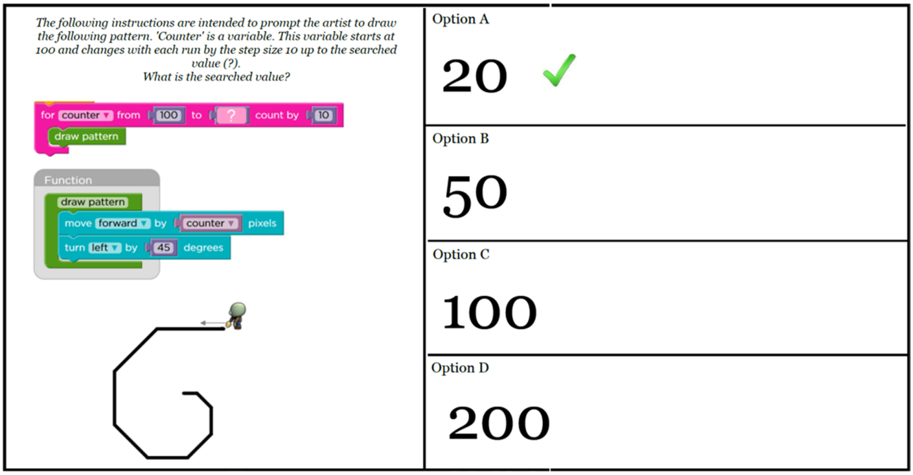


Fig. 4 Item 28, containing sequencing, functions, and variables.

excluded item 1 (see Appendix 1) in almost all cases. Of the students in the sample, 29.7% are on level I, i.e., they are able to systematically solve sequencing tasks.

Level II ($-0.75 \leq \text{Logit} < 0.25$): Level II comprises tasks that include conditionals and/or functions. These are core elements of CT and important in many domains. For instance, they are necessary for an effective use of spreadsheet software like Microsoft Excel. Item 8 is an anchor item for this level (see Fig. 3). In comparison to item 6, the increase in difficulty may be attributed to the use of conditionals. Of the students in the sample, 22.8% are on level II.

Level III ($0.25 \leq \text{Logit}$): On level III, items typically include the use of variables, which is also a core concept of CT (and programming). Item 28 is an anchor for level III (see Fig. 4). Of the students in the sample, 40.1% are on level III; hence, they are systematically able to cope with sequencing, conditionals, functions, and variables.

We cannot set a fourth level because we do not have meaningful item characteristics that justify building such a level, i.e., no suitable anchor items are available.

4.2 RQ2: CT Latent Profiles

4.2.1 Psychometric Validity of the CTS (German Version)

In our sample, the fit-values of the initial version of the CTS with 29 questions indicated room for improvement: YB-$\chi^2(340)=657$ ($p<0.001$), CFI = 0.881, TLI = 0.868, RMSEA = 0.073, SRMR = 0.095. The reasons are mainly due to cross loadings. For instance, the first question of algorithmic thinking also loads significantly on critical thinking and creativity. Discriminant validity is not ensured. Based on a content review, we selected three items for each of the five dimensions. This approach yielded a decent fit: YB-χ^2 $(80)=85$ ($p=0.341$), CFI = 0.997, TLI = 0.996, RMSEA = 0.018 (90% CI [0.000, 0.047]), SRMR = 0.040. Convergent and discriminant validity are fulfilled. The average variance extracted is greater than 0.543 for all five constructs. The heterotrait–monotrait ratio is smaller than 0.706 for all combination of constructs. The five dimensions are

Table 2 Item difficulty, fit, and DIF-effects of CTt (N = 202)

	Item difficulty and discrimination						DIF in Logit				
# German version	# Original version	θ	s.e. θ	wMNSQ	Pt.bis	P+ (%)	Ability	Gender	Age	CL	Prog
2	4	−2.42	0.23	0.99	0.34	87	−0.04	−0.09	−0.01	−0.13	0.40
3	7	−2.75	0.25	0.99	0.30	90	−0.06	0.41	0.18	−0.30	0.40
4	8	−0.32	0.16	1.14	0.42	55	−0.37	0.18	−0.17	0.07	−0.14
5	10	−0.45	0.16	1.08	0.45	58	−0.20	−0.13	0.03	−0.04	−0.11
6	11	−2.22	0.21	1.07	0.33	85	−0.07	0.17	0.09	−0.09	−0.02
7	12	−0.18	0.16	0.97	0.54	53	0.23	0.15	0.21	−0.25	−0.10
8	13	−1.24	0.18	1.15	0.34	72	−0.45	−0.37	−0.01	−0.06	−0.17
9	14	−1.63	0.19	0.97	0.44	78	−0.11	0.10	0.26	−0.41	0.48
11	16	−0.72	0.17	0.92	0.56	63	0.33	0.43	0.02	−0.06	0.58
12	17	−0.80	0.17	1.09	0.41	64	−0.23	−0.22	−0.10	−0.07	−0.47
13	18	−1.40	0.18	1.05	0.39	74	−0.25	−0.37	0.41	−0.25	−0.46
14	19	−0.94	0.17	1.07	0.42	67	−0.23	−0.03	0.03	0.17	−0.23
15	20	−2.01	0.20	0.91	0.45	83	0.38	0.10	−0.04	0.32	−0.01
16	21	−1.53	0.18	0.98	0.46	76	0.05	0.35	−0.33	0.21	0.17
17	22	−0.18	0.16	0.94	0.57	53	0.05	0.26	−0.16	0.18	−0.26
18	23	1.45	0.19	1.22	0.30	25	−0.58	−0.13	0.33	0.08	−0.27
19	24	−1.85	0.20	0.92	0.47	81	0.16	0.03	−0.03	0.16	0.21
20	25	−0.69	0.17	0.89	0.58	62	0.51	0.32	−0.06	0.02	0.03
21	26	−0.94	0.17	0.94	0.53	67	0.38	−0.25	−0.11	0.26	−0.30
22	27	−0.91	0.17	1.04	0.46	66	0.04	−0.17	0.04	0.00	−0.48
23	28	−1.40	0.18	1.01	0.45	74	−0.18	−0.31	−0.18	0.02	−0.06
24	–	−0.50	0.16	0.89	0.59	59	0.14	−0.12	−0.19	0.33	0.14
25	–	−1.46	0.18	0.91	0.52	75	0.29	−0.29	−0.18	0.08	0.21
26	–	−0.26	0.16	0.93	0.57	54	0.20	−0.03	0.12	−0.17	0.05
27	–	0.51	0.17	0.96	0.56	40	0.00	0.05	−0.12	0.09	0.02

Table 2 (continued)

	Item difficulty and discrimination						DIF in Logit				
# German version	# Original version	θ	s.e. θ	wMNSQ	Pt.bis	P+ (%)	Ability	Gender	Age	CL	Prog
28	–	−0.10	0.16	0.97	0.55	51	0.03	−0.01	−0.02	−0.16	0.41

θ item difficulty in Logit, *s.e.* standard error of item difficulty, *wMNSQ* weighted mean square error, *Pt.bis*. point biserial correlation, *P*+percentage of correct responses, *Ability* median of CTt raw score, *CL* computer literacy (INCOBI-R), *Prog*. programming ability (yes/no)

Table 3 Characteristics of used CTS items and constructs (N = 202)

Construct		Item	Mean (SD)	λ	α	ω	AVE	Latent correlations below diagonal, square root of AVE on diagonal, manifest correlations above diagonal				
								(1)	(2)	(3)	(4)	(5)
(1)	Creativity	cr_3	5.6 (1.5)	0.75	0.87	0.87	0.75	0.87	**0.28**	**0.19**	**0.59**	0.17
		cr_4	5.6 (1.3)	0.86								
		cr_5	5.3 (1.3)	0.89								
(2)	Algorithmic thinking	al_3	4.0 (1.9)	0.90	0.90	0.90	0.73	**0.32**	0.85	0.13	**0.53**	−0.00
		al_4	3.7 (1.8)	0.86								
		al_5	4.0 (1.8)	0.83								
(3)	Cooperativity	co_1	4.6 (1.8)	0.85	0.89	0.89	0.70	**0.22**	0.15	0.84	**0.19**	−0.14
		co_2	4.2 (1.8)	0.90								
		co_3	4.8 (1.7)	0.82								
(4)	Critical thinking	cr_1	5.1 (1.4)	0.74	0.80	0.80	0.57	**0.71**	**0.62**	**0.22**	0.75	0.13
		cr_2	4.9 (1.4)	0.75								
		cr_3	4.6 (1.4)	0.78								
(5)	Problem solving	pr_1	5.5 (1.5)	0.79	0.78	0.78	0.54	0.20	−0.03	−0.20	0.15	0.73
		pr_2	5.4 (1.7)	0.68								
		pr_4	5.1 (1.5)	0.74								

Items measured on a 7-point rating scale. λ=standardized factor loading, α Cronbach's alpha, ω Revell's omega total, *AVE* average variance extracted. Figures in bold indicate significant correlations at the 5% level

Table 4 Information criteria, entropies and BLRT results for one to five latent profiles

No. profiles	-LL	No. Par	AIC	CAIC	BIC	aBIC	Entropy	p(BLRT)
1	1421	8	2859	2893	2885	2859	1.000	–
2	1360	13	2747	2803	2790	2749	0.734	0.000
3	1334	18	2704	2782	2764	2707	0.744	0.000
4	1321	23	2687	2786	2763	2690	0.763	0.000
5	1314	28	2680	2800	2772	2684	0.786	0.286

-LL -Log-likelihood, *No. Par.* number of estimated parameters, *AIC* Akaike's information criterion, *CAIC* Consistent AIC, *BIC* Bayesian Information Criterion, *aBIC* adjusted BIC, *BLRT* bootstrap likelihood ratio test

reliably measured (α and ω > 0.77). The characteristics of the refined version of the CTS can be found in Table 3; the questions can be seen in Appendix 2.

Full uniqueness measurement invariance is ensured for the four considered subgroups: gender: $\Delta\chi^2$ (105) = 121.84, p = 0.125; age: $\Delta\chi^2$ (105) = 116.98, p = 0.200; computer literacy: $\Delta\chi^2$ (105) = 122.92, p = 0.112; and ability to program: $\Delta\chi^2$ (105) = 128.94, p = 0.056. Hence, it may be justified to use manifest means for the LPA.

4.2.2 Latent Profiles

The descriptive statistics in Table 3 show that, on average, the students assess their creative thinking, algorithmic thinking, cooperativity, critical thinking, and problem solving above the neutral scale mean (= 4). The highest latent correlation appears between creativity and critical thinking (ρ = 0.71, p < 0.001). However, there are also small and statistically insignificant correlations, e.g., between algorithmic thinking and cooperativity (ρ = 0.15, p = 0.079).

When identifying latent profiles, the dimension *problem solving* was problematic. The likely reason is the reverse coding of the corresponding items; see Appendix 2. Including this dimension in the LPA yielded spurious profiles, e.g., a profile with students who score very low in problem solving and very high in all other dimensions. Since considering this construct could bias the findings, we removed it from the further analysis. This may also be suitable from a conceptual point of view. Creativity, algorithmic thinking, critical thinking, and cooperativity might all be necessary during the course of problem solving; problem solving might be a construct of a different nature.

Table 4 depicts the information criteria, the BLRT results, and the entropies. Based on the information criteria AIC and aBIC, a five-profile solution would be optimal; CAIC, BIC, and BLRT, as well as the analytic hierarchy process of Akogul and Erisoglu (2017), point to four profiles. Moreover, the entropy for this solution is sufficiently high (0.763). A replication of the LPA with 100 bootstrap samples of 150 students lends support to a four-class solution. In 2% of the cases, three profiles are optimal; in 30% of the cases, four profiles; and in 68% of the cases, five profiles. An inspection of the five-class solutions, however, revealed problematic profiles, e.g., profiles with only two students. Since four profiles may be the maximum from a conceptual point of view and three or less profiles seem to be insufficient, four latent profiles could be the appropriate number.

The MANOVA yielded significant different means among the four profiles: F (12, 516) = 69.90, Wilk's Λ = 0.078. p < 0.001, η^2 = 0.573. In other words, 57.3% of the variance

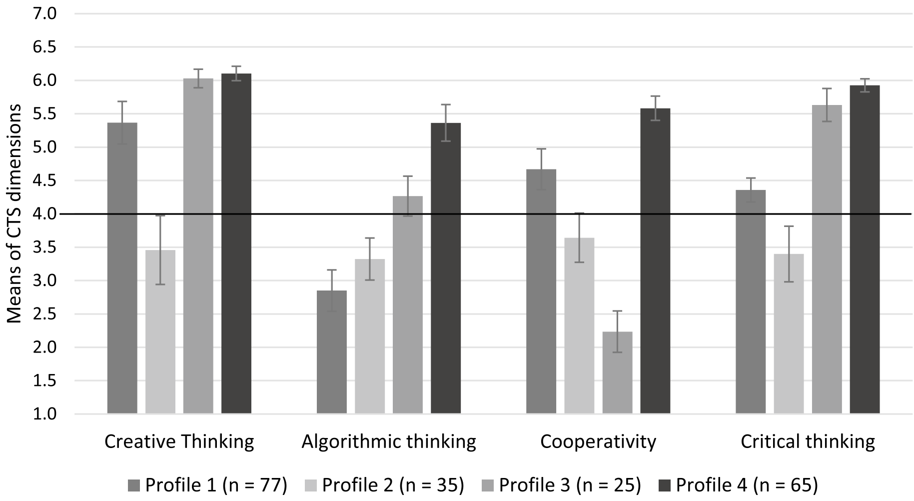


Fig. 5 Description of the four latent profiles based on CTS (N = 202)

can be explained by profile membership. These findings and Fig. 5 point to sufficiently distinct profiles.

The identified profiles may be meaningful from a conceptual point of view:

- Profile 1 (creative thinkers with a focus on collaboration): This profile comprises 77 (38.1%) students and is the largest one. Students in this profile perceive their creative thinking to be well above the neutral scale mean. Moreover, they also assess their cooperativity and critical thinking, on average, as above the neutral scale mean. However, their perceived algorithmic thinking skills are rather low.
- Profile 2 (low-level computational thinkers): These 35 (17.3%) students, on average, self-assess all dimensions of CT as below the scale mean, i.e., they consider themselves as not being capable of performing sufficiently well in the four CT dimensions.
- Profile 3 (computational thinkers with low cooperativity): This profile consists of 25 (12.4%) students, which is the smallest profile. Overall, the students self-assess their creative and critical thinking as very high. Their algorithmic thinking is, on average, slightly above the scale mean. However, these students report very low cooperativity.
- Profile 4 (high-level computational thinkers): Sixty-five (32.2%) students belong to this profile and report high levels across all four CT dimensions.

4.3 RQ3: Relationship Between CTt and CTS Results

The latent correlations between the CTt and the five dimensions of the CTS are with *creativity* 0.271 ($p = 0.002$), *algorithmic thinking* 0.309 ($p < 0.001$), *cooperativity* -0.003 ($p = 0.956$), *critical thinking* 0.408 ($p < 0.001$), and *problem solving* 0.154 (p = 0.085). Considering all CTS dimensions as independent variables, and the CTt as a dependent variable in a latent regression, only *algorithmic thinking* is statistically significant (b = 0.319 Logit, $p < 0.001$). Figure 6 presents the relationship between the three proficiency levels (CTt) and the four latent profiles (CTS). The notches in the boxplots represent the 95% confidence intervals. As can be seen, students in profiles 1 and 2 score significantly lower

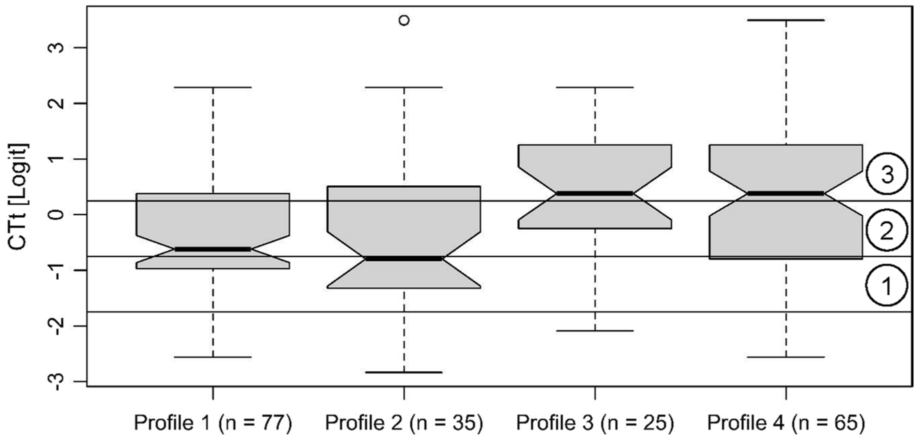


Fig. 6 CTt proficiency in Logit and proficiency levels (1, 2, 3) grouped by latent profile membership (CTS). Notches indicate 95% confidence intervals

in the CTt than students in profiles 3 and 4. The median proficiency in profiles 1 and 2 is around the threshold of proficiency level II, and in profiles 3 and 4, around the threshold of proficiency level III.

5 Discussion, Limitations, and Outlook

5.1 Discussion

The aim of this paper was to contribute to more vivid interpretation of CT assessment findings. To this end, we relied on a performance test (CTt) and a self-assessment instrument (CTS). Both are freely available and internationally accepted standardized instruments (Shute et al., 2017).

In terms of the CTt (RQ1), the main advantage is its reliance on the well-established framework of Brennan and Resnick (2012). Furthermore, the CTt details the cognitive operations that underlie the test items. This may be the basis for a theoretically founded interpretation of the test results. Utilizing IRT, we were able to demonstrate Rasch scalability, which implies the specific objectivity of the test. This is the basis for a criterion-referenced test interpretation. The presented proficiency level model for CT is an additional step in capturing the construct of CT: it adds a model for describing item difficulty and cognitive demand. A proficiency level model could help to communicate the test results. A statement like ‘Student A is able to solve tasks that contain sequencing and conditionals’ is more informative than ‘Student A solved 16 out of 26 items correctly’. It could also help teachers to better tailor their instruction to specific student needs. For instance, a student on proficiency level II might not be able to correctly solve items with variables, but is able to do so as regards items with conditionals and functions. This insight can be utilized for scaffolding processes. Before designing instructional measures, it may be important to evaluate students’ prior knowledge (Bransford et al., 2000). The CTt can carry out this purpose as it does not require specific prior knowledge, such as a programming language.

Concerning the CTS (RQ2), our confirmatory factor analysis indicated, in line with Korkmaz et al. (2017), that five dimensions can be identified. Due to substantial cross-loadings, we had to exclude items from the original version, which contained 29 items. However, the author of this initial version also reduced the initial scale to 20 items in a later study (Korkmaz & Bai, 2019). By means of an LPA, we identified four profiles.

Concerning specific strengths and weaknesses, profiles 1 and 3 are of special interest. Students in profile 1 report a weakness in algorithmic thinking. Algorithmic thinking can be regarded as the link between CT and computer science (Doleck et al., 2017) and may be a technical skill. To foster students' algorithmic thinking skills, it may be preferable to rely on visual programming languages because professional programming languages can be regarded as difficult or boring (Lye & Koh, 2014; Repenning, 2017). Scratch could be a suitable option for this purpose. Grover et al. (2015) demonstrated the effectiveness of Scratch in fostering algorithmic thinking. Moreover, scalable game design (Repenning, 2018) could be a viable option to improve students' algorithmic thinking. The advantage is that students start with a project instead of first being confronted with code and syntax. Students can bring in their creative and critical thinking skills and afterwards learn technical aspects, which keeps them in the zone of proximal flow (Repenning et al., 2015). Besides this, teachers could form groups where students in profile 1 collaboratively learn with students from other profiles, especially students in profile 4 who have high levels of perceived algorithmic thinking skills. Students in profile 1 can bring in their creativity and critical thinking and benefit from the other students' high algorithmic thinking skills.

Profile 3 is remarkable because these students report low cooperativity and, at the same time, high creative and critical thinking. Colloquially, these students might be referred to as nerds. For these students it could be important to increase their cooperativity, which is regarded as a core 21st-century skill (van Laar et al., 2017). Again, Scratch could be a promising approach because it relies on a social learning paradigm and allows members of the community to learn from each other through the opportunity to share and extend projects (Jiang et al., 2021; Repenning et al., 2015; Resnick & Rusk, 2020; Shute et al., 2017). Students in profile 3 could become aware of the benefits of collaboration when working on Scratch projects. Teachers may put a special focus on students in this profile in order to integrate them into the group and facilitate teamwork.

Overall, 67.8% of the students in our sample might require specific support. Students in CT profile 1 may need help with increasing their algorithmic thinking skills. Students in profile 2 score low in all four CTS dimensions. Students in profile 3 seem to have a deficit in cooperativity. Only students in profile 4 report high values among all four CTS dimensions. Teachers could make use of these students to support fellow students with deficits in one or more CT dimension. In general, collaboration seems to be conducive for fostering CT (Denner et al., 2014).

Concerning the relationship between the CTt and CTS results (RQ3), the correlations found in our study are well in line with the correlations reported by Polat et al. (2021), indicating that the findings might be robust across various populations. Moreover, we show that when considering all CTS dimensions as independent variables and the CTt as the dependent variable in a latent regression, only algorithmic thinking is a statistically significant predictor. Hence, the CTt may primarily be related with algorithmic thinking of the ISTE framework. This can also be seen from Figs. 5 and 6. Students in profiles 3 and 4, with the highest reported algorithmic thinking, achieve significantly higher CTt scores than students in profiles 1 and 2, and reach higher proficiency levels.

Since the CTt and CTS rely on different methods, i.e., performance test and self-assessment, it is likely that constructs of a different nature are captured. Self-assessments might capture self-efficacy beliefs (Scherer et al., 2017), which play an important role in predicting (intended) behavior (Fishbein & Ajzen, 2010). For instance, students' perceived cooperativity may be a good predictor for their actual collaboration in CT projects. Moreover, self-assessments are very cost-efficient (Scherer et al., 2017). The CTS requires less than five minutes of test time. Concerning the identification of latent profiles, self-assessment instruments seem to be an established method (e.g., Scherer et al., 2021). Overall, self-assessments may not be inferior to performance tests.

If the purpose is to investigate the nomological net of CT, however, performance tests may be more suitable than self-assessments. Polat et al. (2021) investigated the relationship of CT with the external variables of mathematics and information technologies course achievement. The correlations are substantially higher when the CTt instead of the CTS is used. Moreover, forming proficiency levels based on self-assessments may not be a viable option.

Overall, the CTt and CTS may well complement each other. If we had only used the CTt, the remarkable CT profiles 1 and 3 would not have been detected; students in these profiles may need specific attention.

5.2 Limitations

Our study is not without limitations. A general limitation is that our sample is narrow in scope as it comprises only students from German-speaking Switzerland and from one type of school (high school). Concerning RQ1 and the use of the CTt, one disadvantage is its reliance on dichotomous constructed response items. It may not be able to capture higher-level computational concepts. The formed proficiency levels can be interpreted in a meaningful way because they are linked to previously specified characteristics, e.g., sequencing. However, for high-achieving students, items that cover more advanced computational concepts would be necessary. We will come back to this point in the outlook section below.

Concerning RQ2 and the use of the CTS, we had to exclude items from the original version due to a lack of discriminant validity. On the one hand, this exclusion contributes to the psychometric validity of the instrument and content validity may still be achieved; on the other hand, however, results based on the full version of the CTS could be difficult to compare with our refined version. Besides this, we did not consider the CTS dimension *problem solving* in the LPA. The reason for this was that all the items that operationalize problem solving are reversely coded; consideration of the dimension in the LPA yielded spurious profiles. Moreover, due to our rather small sample size, we had to use a parsimonious LPA model where equal variances across the profiles and covariances of zero are assumed. These assumptions are restrictive (Scherer et al., 2021). Moreover, we used manifest means as the basis for the LPA. This may be unavoidable due to our sample size but it neglects measurement error (Meyer & Morin, 2016). As a robustness check, we also used regression factor scores instead of manifest means; it yielded similar results (see Appendix 3). Against the backdrop of our cross-sectional sample of high school students, it is doubtful whether the identified four profiles can be replicated in samples with younger students and be consistent across time (Meyer & Morin, 2016). Moreover, the identification of the profiles was exploratory, which may be inherent to the LPA but not ideal from a theoretical point of view (Hofmans et al., 2020). Based on our simulation approach, however, we may conclude that four profiles could be a reasonable minimum.

5.3 Outlook

The usefulness of the CTt for assessing students with a CT proficiency equal or higher to those in our sample could be increased. To this end, it would be necessary to set further meaningful proficiency levels. In order to anchor these proficiency levels, items of greater difficulty should be constructed on a criterion basis. Such a criterion could be the CT concept *diffusion* (Repenning, 2017, pp. 18–19). Diffusion goes beyond conditionals, functions, and variables. It is a kind of artificial intelligence and can, for instance, be used to move objects dynamically. As can be seen from the tasks depicted in Figs. 2 and 3, instructions are required to lead Pacman to the ghost. However, if Pacman should autonomously find the ghost, utilizing diffusion is a viable option (see Repenning, 2006). It can be claimed that the ghost has a 'scent' that spreads in the maze. This process is modeled by means of diffusion equations. Pacman can then use a hill-climbing approach to find the ghost. It checks the scent concentration in all four neighboring fields and moves to the field with the highest concentration. This is an efficient approach because it takes obstacles like walls into account. To test students' understanding of this concept, they could be asked how long it will take, depending on the speed of diffusion, to detect the ghost. They may also predict the path Pacman will choose to reach the ghost or implement the hill-climbing approach using visual code blocks. We think including diffusion into the CTt is promising because it is an important concept in many domains (Repenning, 2006). Examples are the diffusion of heat (physics) or osmosis (biology).

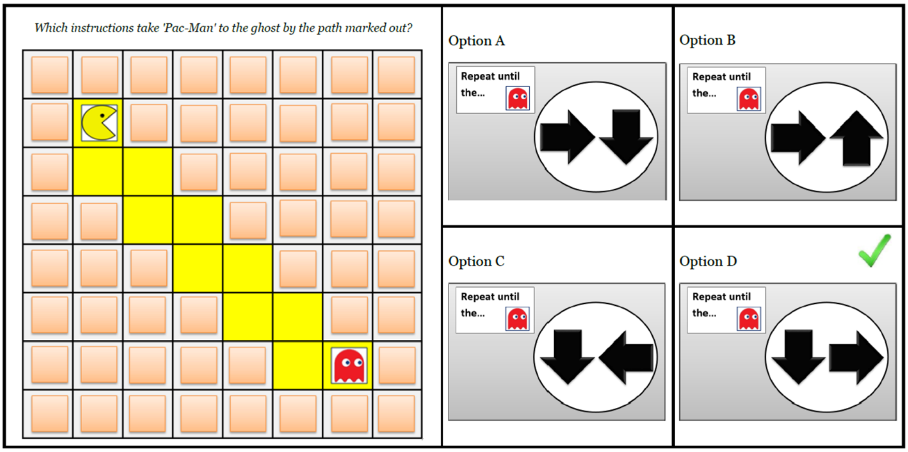
In future research, it could be promising to use performance tests to capture the five dimensions covered by the CTS and form proficiency levels. Suitable tests are available and could be adapted for CT: for creativity (Israel-Fishelson & Hershkovitz, 2022, Appendix C), for algorithmic thinking (Román-González et al., 2017), for cooperativity (Salas et al., 2017), for (complex) problem solving (Greiff et al., 2013), for collaborative problem solving (Stadler et al., 2020), and for critical thinking (Ennis, 1993). If a self-assessment instrument should be used, the computing attitudes survey (Dorn & Tew, 2015) could be a viable alternative to the CTS.

6 Conclusion

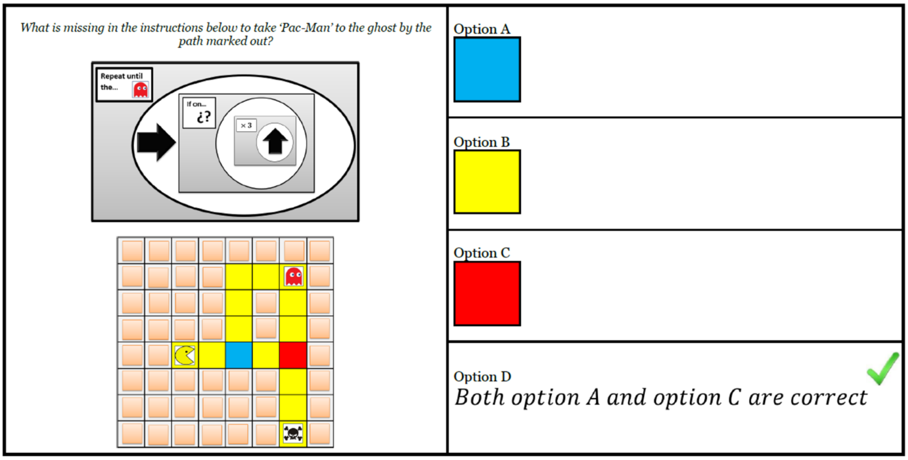
This paper contributes to a better understanding of computational thinking as a construct. Based on a proficiency level model, distinct levels of computational thinking performance can be described, i.e., characteristics of computational thinking tasks that students on a specific level are systematically able to master but which cannot be mastered by students on a lower level. We formed proficiency levels based on the Computational Thinking Test, which covers the cognitive operations (computational concepts) of sequencing, conditionals, functions, and variables. Moreover, we identified latent profiles based on the Computational Thinking Scale. Our findings indicated that 67.8% of the students may need specific guidance and support. Students in the first profile (38.1%) reported deficits in algorithmic thinking while students in the second profile (17.3%) self-assessed themselves as low in all four dimensions: creative thinking, algorithmic thinking, cooperativity, and critical thinking. The third profile (12.4%) comprises students who were reluctant to cooperate. Knowledge about the identified proficiency levels and the four distinctive computational thinking profiles could help teachers offer person-centered guidance and support to their students.

Appendix 1: Excluded items due to psychometric validations

Item 1. Excluded because too easy for the test takers:



Item 10. Excluded because test takers were confused by option D:



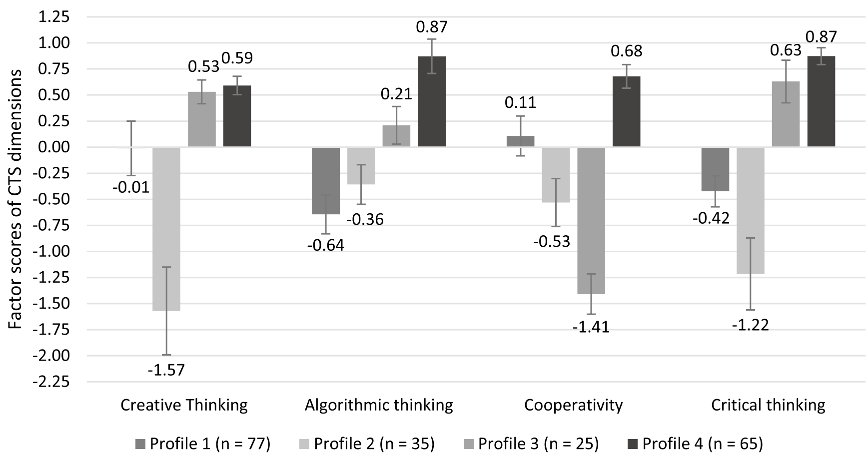
Appendix 2: Refined version of CTS

Creativity	cr_3	"I believe that I can solve most of the problems I face if I have a sufficient amount of time and if I show effort."
	cr_4	"I believe that I can solve possible problems that may occur when I encounter a new situation."

	cr_5	"I trust that I can apply a plan, at the same time as making it, in order to solve a problem."
Algorithmic thinking	al_3	"I think that I am better able to learn instructions with the help of mathematical symbols and concepts."
	al_4	"I can mathematically express the solutions for the problems I face in daily life."
	al_5	"I can digitize a mathematical problem expressed verbally."
Cooperativity	co_1	"I like experiencing cooperative learning together in my group of friends."
	co_2	"In cooperative learning, I think that I attain/will attain more successful results because I am working in a group."
	co_3	"I like solving problems related to a group project together with my friends in cooperative learning."
Critical thinking	cr_1	"I am willing to learn challenging things."
	cr_2	"I am proud of being able to think with great precision."
	cr_3	"I make use of a systematic method while comparing the options at hand and while reaching a decision."
Problem solving	pr_1	"I have problems in demonstrating the solution to a problem in my mind." (R)
	pr_2	"I have difficulties regarding the issue of where and how I should use variables such as X and Y in the solution of a problem." (R)
	pr_4	"I cannot apply the solutions I plan respectively and gradually." (R)

Selection of 15 out of 29 items (Korkmaz et al., 2017, p. 565). Measured on a 7-point rating scale ranging from 'not true at all' to 'entirely true'. R = reverse coding

Appendix 3



Note. Latent profiles based on factor scores (M = 0, SD = 1) are provided to facilitate comparison across studies.

Funding The first author received a postdoctoral fellowship for carrying out this research from the Basic Research Fund at the University of St.Gallen (no. 1031542).

Data availability Readers who wish to examine the data may contact the corresponding author.

Declarations

Conflict of interest No known conflicts of interest associated with this manuscript.

Ethical standards The procedures performed in this study follow the standards of the Ethics Committee of the University of St.Gallen.

References

AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association. https://eric.ed.gov/?id=ED565876

Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy, 19*(9), 452. https://doi.org/10.3390/e19090452

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*(1), 123–140. https://doi.org/10.1007/BF02291180

Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. Learning & Leading with Technology, 38(6), 20–23. https://eric.ed.gov/?id=EJ918910

Brackmann, C. P., Román-González, M., Robles, G., Moreno-León, J., Casali, A., & Barone, D. (2017). Development of computational thinking skills through unplugged activities in primary school. In E. Barendsen (Ed.), Proceedings of the 12th Workshop on Primary and Secondary Computing Education (pp. 65–72). ACM. https://doi.org/10.1145/3137065.3137069

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). How people learn: Brain, mind, experience, and school (Expanded edition). National Academy Press. https://doi.org/10.17226/9853

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. American Educational Research Association Meeting, Vancouver, BC, Canada, 1–25. https://www.media.mit.edu/publications/new-frameworks-for-studying-and-assessing-the-development-of-computational-thinking/

Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion [Introduction to test and questionnaire construction] (3rd ed.). Pearson Studium. https://d-nb.info/1007197234

Buitrago Flórez, F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., & Danies, G. (2017). Changing a generation's way of thinking: Teaching computational thinking through programming. *Review of Educational Research, 87*(4), 834–860. https://doi.org/10.3102/0034654317710096

Carretero, S., Vuorikari, R., & Punie, Y. (2017). DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use. European Union (EU). https://doi.org/10.2760/38842

Chan, S.-W., Looi, C.-K., & Sumintono, B. (2020). Assessing computational thinking abilities among Singapore secondary students: A Rasch model measurement analysis. *Journal of Computers in Education, 8*, 213–236. https://doi.org/10.1007/s40692-020-00177-2

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education, 109*, 162–175. https://doi.org/10.1016/j.compedu.2017.03.001

Denner, J., Werner, L., Campe, S., & Ortiz, E. (2014). Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education, 46*(3), 277–296. https://doi.org/10.1080/15391523.2014.888272

Doleck, T., Bazelais, P., Lemay, D. J., Saxena, A., & Basnet, R. B. (2017). Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: Exploring the relationship between computational thinking skills and academic performance. *Journal of Computers in Education, 4*(4), 355–369. https://doi.org/10.1007/s40692-017-0090-9

Dorn, B., & Tew, A. E. (2015). Empirical validation and application of the computing attitudes survey. *Computer Science Education, 25*(1), 1–36. https://doi.org/10.1080/08993408.2015.1014142

Durak, H. Y., & Saritepeci, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education, 116*, 191–202. https://doi.org/10.1016/j.compedu.2017.09.004

Embretson, S. E., & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. Psychology Science Quarterly, 50(3), 328–344. https://www.semanticscholar.org/paper/Understanding-and-Quantifying-Cognitive-Complexity-Embretson-Daniel/29a75e2f74f0d5a870a846660c617a2273b73975

Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice, 32*(3), 179–186. https://doi.org/10.1080/00405849309543594

Ericson, B., & McKlin, T. (2012). Effective and sustainable computing summer camps. In L. Smith King, D. R. Musicant, T. Camp, & P. Tymann (Eds.), Proceedings of the 43rd ACM technical symposium on Computer Science Education - SIGCSE '12 (p. 289). ACM Press. https://doi.org/10.1145/2157136.2157223

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*(6), 359–374. https://doi.org/10.1016/0001-6918(73)90003-6

Fishbein, M., & Ajzen, I. (2010). Predicting and changing behavior: The reasoned action approach. Taylor and Francis. https://psycnet.apa.org/record/2009-17267-000

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 international report. International Association for the Evaluation of Educational Achievement (IEA). https://www.iea.nl/publications/study-reports/preparing-life-digital-world

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). IEA International Computer and Information Literacy Study 2018: Technical Report. International Association for the Evaluation of Educational Achievement (IEA). https://www.iea.nl/publications/technical-reports/icils-2018-technical-report

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364–379. https://doi.org/10.1037/a0031856

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Grover, S., Pea, R., & Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education, 25*(2), 199–237. https://doi.org/10.1080/08993408.2015.1033142

Grover, S., Pea, R., & Cooper, S. (2016). Factors influencing computer science learning in middle school. In C. Alphonce, J. Tims, M. Caspersen, & S. Edwards (Eds.), Proceedings of the 47th ACM Technical Symposium on Computing Science Education (pp. 552–557). ACM. https://doi.org/10.1145/2839509.2844564

Guggemos, J. (2021). On the predictors of computational thinking and its growth at the high-school level. *Computers & Education*. https://doi.org/10.1016/j.compedu.2020.104060

Guggemos, J., Seufert, S., & Román-González, M. (2019). Measuring computational thinking - Adapting a performance test and a self-assessment instrument for German-speaking countries. *Proceedings of the 16th International Conference Cognition and Exploratory Learning in the Digital Age (CELDA)* (pp. 183–191). https://doi.org/10.33965/celda2019_201911L023

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

Harteis, C., Goller, M., & Caruso, C. (2020). Conceptual change in the face of digitalization: Challenges for workplaces and workplace learning. *Frontiers in Education*. https://doi.org/10.3389/feduc.2020.00001

Hartig, J., & Frey, A. (2013). Sind modelle der item-response-theorie (IRT) das „Mittel der Wahl" für die Modellierung von Kompetenzen? [Benefits and limitations of modeling competencies by means of Item Response Theory (IRT)]. *Zeitschrift Für Erziehungswissenschaft, 16*(S1), 47–51. https://doi.org/10.1007/s11618-013-0386-0

Hartig, J., Frey, A., Nold, G., & Klieme, E. (2012). An application of explanatory item response modeling for model-based proficiency scaling. *Educational and Psychological Measurement, 72*(4), 665–686. https://doi.org/10.1177/0013164411430707

Hofmans, J., Wille, B., & Schreurs, B. (2020). Person-centered methods in vocational research. *Journal of Vocational Behavior, 118*, 103398. https://doi.org/10.1016/j.jvb.2020.103398

Hooshyar, D., Malva, L., Yang, Y., Pedaste, M., Wang, M., & Lim, H. (2021). An adaptive educational computer game: Effects on students' knowledge and learning attitude in computational thinking. *Computers in Human Behavior, 114*, 106575. https://doi.org/10.1016/j.chb.2020.106575

Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296–310. https://doi.org/10.1016/j.compedu.2018.07.004

Ifenthaler, D., Hofhues, S., Egloffstein, M., & Helbig, C. (Eds.). (2021). Digital Transformation of Learning Organizations. Springer. https://doi.org/10.1007/978-3-030-55878-9

Ilic, U., Haseski, H. I., & Tugtekin, U. (2018). Publication trends over 10 years of computational thinking research. *Contemporary Educational Technology., 9*(2), 131–153. https://doi.org/10.30935/cet.414798

Israel, M., Pearson, J. N., Tapia, T., Wherfel, Q. M., & Reese, G. (2015). Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis. *Computers & Education, 82*, 263–279. https://doi.org/10.1016/j.compedu.2014.11.022

Israel-Fishelson, R., & Hershkovitz, A. (2022). Studying interrelations of computational thinking and creativity: A scoping review (2011–2020). *Computers & Education, 176*, 104353. https://doi.org/10.1016/j.compedu.2021.104353

ISTE. (2015). Computational thinking: leadership toolkit. https://www.iste.org/computational-thinking

Jiang, B., Zhao, W., Gu, X., & Yin, C. (2021). Understanding the relationship between computational thinking and computational participation: A case study from Scratch online community. *Educational Technology Research and Development, 69*(5), 2399–2421. https://doi.org/10.1007/s11423-021-10021-8

Kirschner, P. A., & Stoyanov, S. (2020). Educating youth for nonexistent/not yet existing professions. *Educational Policy, 34*(3), 477–517. https://doi.org/10.1177/0895904818802086

Köhler, C., & Hartig, J. (2017). Practical significance of item misfit in educational assessments. *Applied Psychological Measurement, 41*(5), 388–400. https://doi.org/10.1177/0146621617692978

Kong, S. C., & Wang, Y. Q. (2021). Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts. *Computers in Human Behavior, 122*, 106836. https://doi.org/10.1016/j.chb.2021.106836

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior, 72*, 558–569. https://doi.org/10.1016/j.chb.2017.01.005

Korkmaz, Ö., & Bai, X. (2019). Adapting computational thinking scale (CTS) for Chinese high school students and their thinking scale skills level. *Participatory Educational Research, 6*(1), 10–26. https://doi.org/10.17275/per.19.2.6.1

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Kong, S. C. (2019). Components and methods of evaluating computational thinking for fostering creative problem-solvers in senior primary school education. In S. C. Kong & H. Abelson (Eds.), Computational thinking education (pp. 119–141). Springer. https://doi.org/10.1007/978-981-13-6528-7_8

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*. https://doi.org/10.5334/irsp.289

Lohr, A., Stadler, M., Schultz-Pernice, F., Chernikova, O., Sailer, M., Fischer, F., & Sailer, M. (2021). On powerpointers, clickerers, and digital pros: Investigating the initiation of digital learning activities by teachers in higher education. *Computers in Human Behavior, 119*, 106715. https://doi.org/10.1016/j.chb.2021.106715

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior, 41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Mair, P., & Hatzinger, R. (2007) Extended Rasch modeling The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20(9), 1–20. https://doi.org/10.18637/jss.v020.i09

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing, 48*(2), 175–186. https://doi.org/10.1111/j.1365-2648.2004.03185.x

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Meyer, J. P., & Morin, A. J. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior, 37*(4), 584–612. https://doi.org/10.1002/job.2085

Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 39–59. https://doi.org/10.1080/10705511.2014.919825

Mühling, A., Ruf, A., & Hubwieser, P. (2015). Design and first results of a psychometric test for measuring basic programming abilities. In A. Mühling, A. Ruf, & P. Hubwieser (Eds.), Design and First Results of a Psychometric Test for Measuring Basic Programming Abilities (pp. 2–10). ACM. https://doi.org/10.1145/2818314.2818320

OECD. (2015). PISA 2012 Technical Report. OECD Publishing. https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

OECD. (2017). PISA 2015 Technical Report. OECD Publishing. http://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59–81. https://doi.org/10.1080/00461520.2016.1145550

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43*(4), 295–312. https://doi.org/10.1111/j.1745-3984.2006.00018.x

Polat, E., Hopcan, S., Kucuk, S., & Sisman, B. (2021). A comprehensive assessment of secondary school students' computational thinking skills. *British Journal of Educational Technology, 52*(5), 1965–1980. https://doi.org/10.1111/bjet.13092

Prenzel, M., Drechsel, B., & Kramer, K. (1998). Lernmotivation im kaufmännischen Unterricht: Die Sicht von Auszubildenden und Lehrkräften [Learning motivation in business education: the perspective of trainees and teachers]. *Zeitschrift für Berufs- Und Wirtschaftspädagogik*, special issue 14, 169–187

Relkin, E., de Ruiter, L. E., & Bers, M. U. (2021). Learning to code and the acquisition of computational thinking by young children. *Computers & Education, 169*, 104222. https://doi.org/10.1016/j.compedu.2021.104222

Repenning, A. (2006). Collaborative diffusion. Companion to the 21st ACM SIGPLAN Symposium on Object-Oriented Programming Systems, Languages, and Applications, 574–585. https://doi.org/10.1145/1176617.1176630

Repenning, A. (2017). Moving beyond syntax: Lessons from 20 years of blocks programing in AgentSheets. *Journal of Visual Languages and Sentient Systems, 3*(1), 68–91. https://doi.org/10.18293/VLSS2017-010

Repenning, A. (2018). Scale or fail: Moving beyond self-selected computer science education in Switzerland. *Communications of the ACM, 61*(5), 40–42. https://doi.org/10.1145/3199603

Repenning, A., Webb, D. C., Koh, K. H., Nickerson, H., Miller, S. B., Brand, C., Her Many Horses, I., Basawapatna, A., Gluck, F., Grover, R., Gutierrez, K., & Repenning, N. (2015). Scalable game design: A strategy to bring systemic computer science education to schools through game design and simulation creation. *ACM Transactions on Computing Education*, *15*(2), 1–31. https://doi.org/10.1145/2700517

Resnick, M., & Rusk, N. (2020). Coding at a crossroads. *Communications of the ACM, 63*(11), 120–127. https://doi.org/10.1145/3375546

Richter, T., Naumann, J., & Horz, H. (2010). A revised version of the computer literacy inventory (INCOBI-R). *Zeitschrift Für Pädagogische Psychologie, 24*(1), 23–37. https://doi.org/10.1024/1010-0652/a000002

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education*. https://doi.org/10.3389/feduc.2020.589965

Robitzsch, A., Kiefer, T., & Wu, M. (2020). Package 'TAM'. https://cran.r-project.org/web/packages/TAM/TAM.pdf

Román-González, M. (2015). Computational thinking test: Design guidlines and content validation. https://doi.org/10.13140/RG.2.1.4203.4329

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior, 72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In S. C. Kong & H. Abelson (Eds.), *Computational thinking education* (pp. 79–98). Springer. https://doi.org/10.1007/978-981-13-6528-7_6

Rose, S. P., Habgood, M. J., & Jay, T. (2019). Using pirate plunder to develop children's abstraction skills in Scratch. In R. Mandryk, M. Hancock, M. Perry, & A. Cox (Eds.), Extended Abstracts of the 2019

CHI Conference on Human Factors in Computing Systems (pp. 1–6). ACM Press. https://doi.org/10.1145/3290607.3312871

Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C., & Schmidt, J. (2018). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software, 3*(30), 978. https://doi.org/10.21105/joss.00978

Rosman, T., Mayer, A.-K., & Krampen, G. (2015). Combining self-assessments and achievement tests in information literacy assessment: Empirical results and recommendations for practice. *Assessment & Evaluation in Higher Education, 40*(5), 740–754. https://doi.org/10.1080/02602938.2014.950554

Rosseel, Y. (2012c). lavaan: An R package for structural equation modeling. *Journal of Statistical Software., 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Salas, E., Reyes, D. L., & Woods, A. L. (2017). The assessment of team performance: observations and needs. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 21–36). Springer. https://doi.org/10.1007/978-3-319-33261-1_2

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review, 63*(1), 34–62. https://doi.org/10.1007/BF03396886

Scherer, R., Howard, S. K., Tondeur, J., & Siddiq, F. (2021). Profiling teachers' readiness for online teaching and learning in higher education: Who's ready? *Computers in Human Behavior, 118*, 106675. https://doi.org/10.1016/j.chb.2020.106675

Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology, 111*(5), 764–792. https://doi.org/10.1037/edu0000314

Scherer, R., Tondeur, J., & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the technological, pedagogical, and content knowledge (TPACK) model. *Computers & Education, 112*, 1–17. https://doi.org/10.1016/j.compedu.2017.04.012

Seiter, L., & Foreman, B. (2013). Modeling the learning progressions of computational thinking of primary grade students. In B. Simon, A. Clear, & Q. Cutts (Eds.), Proceedings of the ninth annual international ACM conference on International computing education research (pp. 59–66). ACM. https://doi.org/10.1145/2493394.2493403

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior, 115*, 106552. https://doi.org/10.1016/j.chb.2020.106552

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Stadler, M., Herborn, K., Mustafić, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks. *Computers & Education, 157*, 103964. https://doi.org/10.1016/j.compedu.2020.103964

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education, 148*, 103798. https://doi.org/10.1016/j.compedu.2019.103798

Tondeur, J., Scherer, R., Baran, E., Siddiq, F., Valtonen, T., & Sointu, E. (2019). Teacher educators as gatekeepers: Preparing the next generation of teachers for technology integration in education. *British Journal of Educational Technology, 50*(3), 1189–1209. https://doi.org/10.1111/bjet.12748

Ulitzsch, E., von Davier, M., & Pohl, S. (2020). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response. *The British Journal of Mathematical and Statistical Psychology, 73*(Suppl 1), 83–112. https://doi.org/10.1111/bmsp.12188

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

van Laar, E., van Deursen, A. J., van Dijk, J. A., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577–588. https://doi.org/10.1016/j.chb.2017.03.010

Vanslambrouck, S., Zhu, C., Pynoo, B., Lombaerts, K., Tondeur, J., & Scherer, R. (2019). A latent profile analysis of adult students' online self-regulation in blended learning environments. *Computers in Human Behavior, 99*, 126–136. https://doi.org/10.1016/j.chb.2019.05.021

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies, 20*(4), 715–728. https://doi.org/10.1007/s10639-015-9412-6

Werner, L., Denner, J., & Campe, S. (2012). The fairy performance assessment: measuring computational thinking in middle school. 43rd ACM Technical Symposium on Computer Science Education, 215–220. https://doi.org/10.1145/2157136.2157200

Weintrop, D., Wise Rutstein, D., Bienkowski, M., & McGee, S. (2021). Assessing computational thinking: An overview of the field. *Computer Science Education, 31*(2), 113–116. https://doi.org/10.1080/08993408.2021.1918380

Wilson, M. (2005). Constructing measures: An item response modeling approach. Psychology Press. https://psycnet.apa.org/record/2004-22396-000

Willse, J. T. (2011). Mixture rasch models with joint maximum likelihood estimation. *Educational and Psychological Measurement, 71*(1), 5–19. https://doi.org/10.1177/0013164410387335

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions Series a, Mathematical, Physical, and Engineering Sciences, 366*(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118

Yadav, A., Hong, H., & Stephenson, C. (2016). Computational thinking for all: Pedagogical approaches to embedding 21st century problem solving in K-12 classrooms. *TechTrends, 60*(6), 565–568. https://doi.org/10.1007/s11528-016-0087-7

Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational thinking in elementary and secondary teacher education. *ACM Transactions on Computing Education, 14*(1), 1–16. https://doi.org/10.1145/2576872

Zhao, W., & Shute, V. J. (2019). Can playing a video game foster computational thinking skills? *Computers & Education, 141*, 103633. https://doi.org/10.1016/j.compedu.2019.103633

Studie 2:

Guggemos, J. (2021). On the predictors of computational thinking and its growth at the high-school level. *Computers & Education*, 161, 104060.

Computers & Education 161 (2021) 104060

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

On the predictors of computational thinking and its growth at the high-school level

Josef Guggemos

School of Management, Institute of Business Education and Educational Management, University of St.Gallen, St. Jakob-Strasse 21, 9000, St. Gallen, Switzerland

ARTICLE INFO

Keywords:
Computational thinking level and growth
Gender and home environment
Reasoning skills
Motivation
Longitudinal study

ABSTRACT

Computational thinking (CT) is a key 21st-century skill. This paper contributes to CT research by investigating CT predictors among upper secondary students in a longitudinal and natural classroom setting. The hypothesized predictors are grouped into three areas: student characteristics, home environment, and learning opportunities. CT is measured with the Computational Thinking Test (CTt), an established performance test. N = 202 high-school students, at three time points over one school year, act as the sample and latent growth curve modeling as the analysis method. CT self-concept, followed by reasoning skills and gender, show the strongest association with the level of CT. Computer literacy, followed by duration of computer use and formal learning opportunities during the school year, have the strongest association with CT growth. Variables from all three areas seem to be important for predicting either CT level or growth. An explained variance of 70.4% for CT level and 61.2% for CT growth might indicate a good trade-off between the comprehensiveness and parsimony of the conceptual framework. The findings contribute to a better understanding of CT as a construct and have implications for CT instruction, e.g., the role of computer science and motivation in CT learning.

1. Introduction

Computational thinking (CT) is regarded as a key 21st-century skill (Voogt, Fisser, Good, Mishra, & Yadav, 2015; Wing, 2006). It may be a valuable resource for solving problems in a wide range of subjects and workplace settings (Buitrago Flórez et al., 2017; Barr, Harrison, & Conery, 2011; Yadav, Good, Voogt, & Fisser, 2018). Already used in the 1980s, the term CT experienced a revival due to Wing's (2006) seminal article, which might be the starting point for the current discussion about CT (Angeli et al., 2016). Wing conceptualized CT as "solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science" (Wing, 2006, p. 33). As a theoretical basis for CT, the framework of Brennan and Resnick (2012) is often utilized. It comprises three dimensions (Brennan & Resnick, 2012): Computational concepts, computational practices, and computational perspectives. Several literature reviews have aimed at identifying core facets of CT (Angeli et al., 2016; Hsu, Chang, & Hung, 2018; Shute, Sun, & Asbell-Clarke, 2017). Such core facets could be: abstraction, decomposition, algorithms, and debugging.

As a construct, CT may still be in its infancy (Shute et al., 2017). To contribute to a better understanding of the process, the purpose of our research is to predict (Newman, Ridenour, Newman, & DeMarco, 2007), i.e., we test the association of CT with variables that might explain the CT level and its development. In this vein, the paper at hand addresses four research desiderata. First, the available studies that aim at predicting CT often focus on selected areas, e.g., family background. Since predictors from different areas are likely

E-mail address: josef.guggemos@unisg.ch.

https://doi.org/10.1016/j.compedu.2020.104060
Received 22 March 2020; Received in revised form 15 October 2020; Accepted 20 October 2020
Available online 23 October 2020

to be interrelated, a study that more comprehensively captures potential predictors may be necessary to identify their unique effects. Second, research on predictors of CT is mainly cross-sectional; longitudinal studies to examine predictors of CT development are missing. The third research desideratum is the investigation of CT learning in naturalistic classroom settings (Ilic, Haseski, & Tugtekin, 2018; Lye & Koh, 2014). Interventions where students voluntarily participate may not accurately represent students in general. Fourth, high-school students are currently rarely studied in comparison to those from elementary and middle schools. However, no rationale suggests that elementary and middle schools represent the only critical stages of CT learning (Tang, Yin, Lin, Hadad, & Zhai, 2020). To address the four research desiderata, the paper at hand aims at answering the research question:

What are the predictors of high-school students' CT level and CT growth on the individual level?

To this end, we develop hypotheses on predictors of the level and growth of CT. As we are focusing on the individual level, the classroom or school context is not considered. We test our hypotheses using a sample of 202 high-school students (11th grade) who we surveyed/tested at three intervals over one school year. A latent growth curve model acts as the method because it allows us to make inferences about the initial level of an outcome variable (CT), as well as its rate of change (Pakpahan, Hoffmann, & Kröger, 2017), and thus to answer our research question.

2. State of research and hypotheses

2.1. Review of available studies

From the available studies, the work of Durak and Saritepeci (2018) is most closely related to our research. The authors utilized a cross-sectional sample of 156 students from grades 5 to 12. They reported educational level (grade), mathematics and science class performances, and ways of thinking as significant predictors of CT. Building on this study, it could be beneficial to also consider students' CT motivation and family background because those factors may be important when investigating CT (Del Olmo-Muñoz, Cózar-Gutiérrez, & González-Calero, 2020; Fraillon, Ainley, Schulz, Friedman, & Duckworth, 2019; Repenning et al., 2015). Moreover, Durak and Saritepeci relied on a self-assessment instrument (Korkmaz, Çakir, & Özden, 2017). CT performance tests and self-assessment instruments tend to capture different constructs (Guggemos, Seufert, & Román-González, 2019; Román-González, Moreno-León, & Robles, 2019). It may be beneficial to complement the research of Durak and Saritepeci (2018) by using a performance test that captures CT skills.

Werner, Denner, and Campe (2012) relied on a cross-sectional sample of 311 middle-school students aged 10 to 14. They reported a positive relationship between CT and pair programming, parental education, the father or mother working with computers, speaking more English (native language) at home, interest in computer science classes, grades in school, frequency of computer use, confidence with computers, and attitudes towards computers. A predictor missing in this comprehensive study may be reasoning skills.

Román-González, Pérez-González, and Jiménez-Fernández (2017) and Román-González, Pérez-González, Moreno-León, and Robles (2018) carried out a criterion validity study of the Computational Thinking Test (CTt), a CT performance test. Their cross-sectional sample comprises 1251 students from grades 5 to 10. They investigated the relationship of the CTt with various cognitive variables, e.g., reasoning, and non-cognitive ones, e.g., self-efficacy. By nature, the work is correlational and students' family background and learning opportunities are not considered.

In the large-scale International Computer and Information Literacy Study 2018 (ICILS), CT was an option in which eight countries participated (Fraillon et al., 2019). Using 8th graders as the cross-sectional sample, the ICILS found significantly higher CT scores among male students in comparison to females, for the overall sample. Moreover, the ICILS reported a positive association of CT with socioeconomic background, access to computers at home, and experience using computers. A negative association is found with an immigration background. Evidence about the association with motivation and cognitive dispositions, like reasoning or mathematical skills, however, is not provided.

In terms of CT development (growth), quantitative studies are mostly carried out using (quasi-)experimental designs where the causal effect of specific interventions on CT learning is of interest (Cutumisu, Adams, & Lu, 2019). However, these intervention studies mostly do not focus on CT learning in naturalistic classroom settings over a substantial period of time.

2.2. Conceptual framework and hypotheses

2.2.1. Conceptual framework

According to the research question, we only consider the individual level. To structure our hypotheses, we rely on the conceptual framework of the ICILS. It distinguishes Antecedents and Processes. "Antecedents are exogenous factors that condition the ways in which (…) learning takes place" (Fraillon et al., 2019, p. 6). Antecedents comprise student variables, e.g., gender, and home environment variables, e.g., parents' socioeconomic status. "Processes are those factors that directly influence (…) learning" (Fraillon et al., 2019, p. 6). Such CT learning opportunities can be formal or informal in nature (Grover & Pea, 2013; Wing, 2008).

2.2.2. Antecedents – student

Gender differences are an important topic in CT research (Shute et al., 2017). Durak and Saritepeci (2018) argued that reasons for gender differences in CT may be due to differences in self-efficacy and interest, which might be attributed to gender stereotypes (Master, Cheryan, & Meltzoff, 2016). Overall, the empirical evidence is inconclusive. Grover, Pea, and Cooper (2015) found higher CT levels for female middle-school students in comparison to males. Werner et al. (2012) did not find any gender differences among middle-school students. Atmatzidou and Demetriadis (2016), based on a sample of junior high and high vocational students, reported that females need more learning time but reach the same CT levels as their male counterparts. The ICILS, across the eight countries in the sample, reported significantly higher CT scores for males in comparison to females. Román-González et al. (2017) used Spanish secondary students aged 10 to 16 as a sample and found an increasing gender CT gap in favor of males as students become older. Since our study is in the realm of high-school education, we expect:

H1. The gender 'female' negatively predicts the level and growth of CT.

CT is conceptualized as a problem-solving methodology across subjects (Barr & Stephenson, 2011; Wing, 2006). In light of this, it may be strongly related with reasoning (Ambrosio, Xavier, & Georges, 2015; Buitrago Flórez et al., 2017; Wüstenberg, Greiff, & Funke, 2012). Reasoning is a core facet of general intelligence. It can be broadly defined as "the process of drawing conclusions" (Leighton, 2009, p. 3). In terms of CT core facets, reasoning may be involved in all of them: abstractions, algorithmic thinking, decomposition, and debugging. The strong conceptual relationship between CT and reasoning is supported by empirical evidence. Román-González et al. (2017) reported a medium correlation of CT with reasoning skills. Hence, we assume:

H2. Reasoning skills positively predict the level and growth of CT.

According to Wing (2006, p. 33), CT may be an analytical skill, like "reading, writing, and arithmetic". In terms of mathematical skills, Wing (2008) claimed that CT and mathematical thinking share the same general way of approaching problems. Sneider, Stephenson, Schafer, and Flick (2014) demonstrated a substantial overlap between CT and mathematical thinking, e.g., 'modeling'. In light of this, we might expect a positive association between mathematical skills and CT. Durak and Saritepeci (2018), as well as Grover et al. (2015) and Grover, Pea, and Cooper (2016), reported higher CT levels for students with higher academic success in mathematics. Concerning language skills, Barr and Stephenson (2011) claimed that CT concepts are also present in the language arts. An example is "Write instructions" as a manifestation of an algorithm (Barr & Stephenson, 2011, p. 52). Zhang and Nouri (2019) showed in their literature review that reading is regularly regarded as a part of CT. Against this background, language skills could be positively associated with CT. Román-González et al. (2018) reported a medium positive correlation between CT and verbal ability. In sum, we hypothesize:

H3. Mathematical skills positively predict the level and growth of CT.

H4. Language skills positively predict the level and growth of CT.

The relationship between programming and CT is often thematized. In line with Wing (2006), Israel, Pearson, Tapia, Wherfel, and Reese (2015) regard the use of computers to model ideas and programming as an integral part of CT. Concepts of programming, e.g., loops and conditionals, are elements of the Brennan and Resnick (2012) CT framework. Buitrago Flórez et al. (2017), as well as Lye and Koh (2014), argue that by means of programming, several core facets of CT can be addressed. Shute et al. (2017) concluded there is a close relationship between CT and programming skills due to similar underlying cognitive processes. Hsu et al. (2018) reported, based on their review of the literature, that programming is widely used to teach CT. Against this background, the ability to program may be positively associated with CT. Concerning empirical evidence, Grover et al. (2016) highlighted the positive influence of programming experience on CT; Scherer, Siddiq, and Sánchez Viveros (2019) concluded, based on a meta-analysis, that CT can be taught through programming. However, Duncan and Bell (2015) did not find a positive influence of teaching programming on CT levels. In sum, we expect:

H5. The ability to program positively predicts the level and growth of CT.

CT can be regarded as part of an overall digital literacy that also includes computer (and information) literacy (Yadav et al., 2018). We define computer literacy as the procedural and declarative knowledge that enables an individual to deal competently with the computer and thus to use information technologies in an individually and socially successful way (Richter, Naumann, & Horz, 2010). Many authors, including Wing (2006), stress that computer literacy is different from CT. However, the question remains whether computer literacy is conducive to CT or not. Since CT aims at representing a problem in such a way that a computer can solve it (Israel et al., 2015; Wing, 2006), knowledge about the capabilities of computers may be beneficial. Moreover, CT is often taught using computers and technology (Hsu et al., 2018). Students with higher computer literacy may benefit more from CT instruction because they are more proficient in using computers as learning tools. Concerning empirical evidence, Yeh et al., 2012 used Microsoft Excel, namely functions, to promote CT and reported positive effects. The ICILS also found a strongly positive correlation between information and computer literacy and CT. Against this background, we hypothesize:

H6. Computer literacy positively predicts the level and growth of CT.

According to the expectancy-value model (EVM) (Wigfield & Eccles, 2000), the expectation of success and subjective task value drive the level of achievement motivation. The expectation of success depends on the person's self-concept, which can be broadly defined as the perception of oneself (Shavelson, Hubner, & Stanton, 1976). Drawing from this, CT self-concept could be defined as the perception about oneself in the field of CT. A core element of self-concept is the perceived competence (Bong & Skaalvik, 2003). As such, it may be closely related to self-efficacy. Indeed, domain specific self-concept and self-efficacy are often hard to separate (Bong & Skaalvik, 2003). The main difference might be the time orientation: self-concept is relatively stable whereas self-efficacy is malleable. In both cases, if a person expected to master CT tasks, they would put more effort into CT learning, overcoming obstacles, and CT activities in general. To capture the expectation of success component of the EVM, we rely on self-concept (Retelsdorf, Köller, & Möller, 2011). For our longitudinal study, in comparison to self-efficacy this more time-stable construct might be more suitable. Empirical evidence is available for self-efficacy. Román-González et al. (2018) and Ketenci, Calandra, Margulieux, and Cohen (2019) reported a positive correlation with CT. Overall, we hypothesize:

H7. CT self-concept positively predicts the level and growth of CT.

The second component of the EVM addresses the perceived task value. Following Wigfield and Eccles (2002), the individual perception of usefulness plays a central role. Students who regard CT as more important for their academic and personal success are expected to put more effort into CT learning. This is also expected from students who enjoy engaging in CT tasks and are interested in them, regardless of external incentives. The described elements of perceived task value are consistent with the self-determination theory (Ryan & Deci, 2000); they may be manifestations of self-determined motivation, i.e., the person engages in the subject independently from external influences. Self-determined motivation is an important predictor for high-level learning processes (Seidel, Rimmele, & Prenzel, 2005). Repenning et al. (2015) stressed the importance of motivation for CT learning. For primary students, Kong, Chiu, and Lai (2018) showed a strong positive association between interest and 'programming empowerment', a construct closely related to CT. Ketenci et al. (2019), however, did not find a significant correlation between interest and CT. Drawing on the EVM, we hypothesize:

H8. 'Self-determined motivation' positively predicts the level and growth of CT.

2.2.3. Antecedents – home environment

The home environment is a regularly used predictor of educational outcome (Rutkowski & Rutkowski, 2013). An important aspect of home environment is 'Socioeconomic and Cultural Status' (SECS), which comprises parental income, parental education, parental occupation, and the availability of cultural goods at home. The rationale is that families with a higher SECS are able and willing to provide more favorable learning environments (Retelsdorf et al., 2011). In terms of empirical evidence, the ICILS consistently reported higher CT scores for students from families with a higher SECS. Werner et al. (2012) reported a positive correlation between parental education and CT performance. We hypothesize:

H9. SECS positively predicts the level and growth of CT.

Another important aspect of home environment might be migration (OECD, 2015). Reasons for the lower performances of students from families with migration background could be due to their concentration in disadvantaged schools and language-related issues. The ICILS reported a significantly lower CT score for students from immigrant families in comparison to those from non-immigrant families. Hence, we hypothesize:

H10. A migration background is negatively associated with the level and growth of CT.

2.2.4. Processes – learning opportunities

Formal and informal learning opportunities may be necessary to foster CT (Grover & Pea, 2013; Wing, 2008). Formal learning can be regarded as learning that takes place in courses at school. In light of this, the extent of formal CT learning opportunities is difficult to capture because CT, by its nature, can be part of (almost) every subject (Webb et al., 2018). Hence, for formal learning opportunities, we hypothesize on a general level:

H11. Formal learning opportunities during the school year positively predict the growth of CT.

Although CT could be part of every subject, it is deeply rooted in computer science education (Grover & Pea, 2013) and draws on basic concepts of computer science (Wing, 2006). Hence, we regard computer science instruction as a formal learning opportunity for CT and hypothesize on a more specific level:

H12. Computer science instruction positively predicts the level and growth of CT.

Besides formal learning opportunities, CT could also be fostered in informal settings, i.e., outside of school courses. Durak and Saritepeci (2018) hypothesized that the use of information and communication technology (ICT) and the internet has a positive influence on CT. They argue that interacting with ICT is important for reflecting on CT, and that the use of ICT may be conducive to CT. However, both hypotheses were rejected. A reason for this could be that students use digital devices like smartphones to a great extent for leisure activities (Fraillon et al., 2019), e.g., listening to music or using social media. These activities might not be conducive to CT. In light of this, the use of computers (PC and laptop) may be a better indicator for informal learning opportunities. Indeed, Werner et al. (2012) reported a positive correlation between frequency of computer use and CT. We therefore hypothesize:

H13. Duration of computer use positively predicts the level and growth of CT.

3. Method

3.1. Data collection and sample

Our sample comprises N = 202 students from the 11th (second last) grade of three 'Gymnasium Helveticum' (high schools) in German-speaking Switzerland who received 90 min of computer science instruction per week during the school year. A synopsis of the CT-related content in the curriculum can be found in Table S1 in the Supplementary Material; CT is explicitly mentioned as a curricular goal. The students are nested in twelve classes. Data was collected using Questback Unipark at the beginning (t1 = August), the middle (t2 = January), and the end (t3 = June) of the school year 2018/19, each within a two-week time slot. Data about CT performance was collected at all three time points, concerning the formal learning opportunities during the school year in t3, and about all other covariates in t1. The class teachers supervised the students and ensured a suitable test environment. The scheduled time was 90 min in t1 and 45 min in both t2 and t3. Teachers allowed all students to finish their work. On average, the students were aged 17.23 years (SD = 0.85 years) at the beginning of the data collection. Further characteristics of the sample can be found in the two last lines of Table 2.

3.2. Measures

3.2.1. Outcome variable – CT

To measure CT skills, we utilized the CTt (Román-González et al., 2017), a widely accepted performance test (Cutumisu et al., 2019; Shute et al., 2017; Zhao & Shute, 2019). It aims at measuring "the ability to formulate and solve problems by relying on the fundamental concepts of computing, and using logic-syntax of programming languages: basic sequences, loops, iteration, conditionals, functions and variables" (Román-González et al., 2017, p. 681). The well-established CT framework of Brennan and Resnick (2012) acts as the theoretical background. Moreover, the CTt covers core facets of CT: abstraction, decomposition, algorithms, and debugging. We adapted the CTt to the high-school level, resulting in a performance test with 26 items. To this end, we selected more difficult items in comparison to the initial version of the CTt from the item pool of Román-González (2015). We validated our version of the CTt and demonstrated Rasch-scalability (Guggemos, Seufert, & Román-González, 2019; a summary of this validation can be found in the Supplementary Material). This also implies test fairness, which means that the test "does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct" (AERA, APA, & NCME, 2014, p. 50). Moreover, we provided evidence for unidimensionality. Our CTt version in general meets the quality criteria applied in the PISA studies (OECD, 2017, pp. 131–134). The weighted mean square error (wMNSQ) lies between 0.80 and 1.20 with the exception of only one item (1.22). This suggests an overall good fit with the Rasch model. The Wright Map (see Fig. S1 in the Supplementary Material) demonstrates an acceptable fit between student ability and item difficulty. All point-biserial correlations are higher than 0.30 revealing a sufficient discriminatory power of the items. The percentage of correct answers for all items lies between 0.90 and 0.25 indicating the absence of too easy or too difficult items. A sample item of our CTt version is depicted in Fig. 1. We used WLE estimators to measure student CT performance (Warm, 1989). The WLE reliability equals 0.81 in t1 and 0.85 in t2 and in t3. For a more vivid presentation, CT scores were transformed to a scale of M = 50 and SD = 10 at t1 (Retelsdorf et al., 2011). At this level of CT, the probability of mastering the item provided in Fig. 1 equals 51%.

3.2.2. Predictors

This section describes the operationalization of the hypothesized CT predictors. The subsequent section addresses their scaling, reliability assessment, and the taken measures.

Table 1
Operationalization of predictors, assessment of reliability, and taken actions.

Predictor	Instrument	Scale	Reliability	Action
Gender	self-report	female = 1, male = 0	n/a	manifest variable
Reasoning skills	6-items performance test	binary, multiple-choice	ω = .76, α = .70 ✓	sum as manifest variable
Mathematical skills	self-report grade	1–6	n/a	manifest variable
Language skills	self-report grade	1–6	n/a	manifest variable
Ability to program	1-item self-assessment	binary	n/a (single-item)	*error variance considered*
Computer literacy	20-item performance test	binary, multiple-choice	ω = .79, α = .76 ✓	mean as manifest variable
CT Self-concept	6-item self-assessment	rating 1–7	ω = .93, α = .88 ✓	mean as manifest variable
Self-determined CT motivation	9-item self-assessment	rating 1–7	ω = .95, α = .93 ✓	mean as manifest variable
Parents' SECS	self-report	3 components: ISEI father, ISEI mother, no. of books at home	n/a (formative measurement)	first principal component as manifest variable
Migration background	self-report	yes = 1, no = 0	n/a	manifest variable
Formal learning in school year	5-item self-assessment	rating 1–7	ω = .87, α = .85 ✓	mean as manifest variable
Past computer science instruction	self-report	weekly number of lessons	n/a	manifest variable
Duration of computer use	self-report	rating 1–7 ranging from 0 – more than 6 h/day	n/a	manifest variable

Note. ω = Revelle's Omega Total, α = Cronbach's Alpha. ✓ = sufficient reliability. Self-report = fact is reported. Self-assessment = (subjective) assessment necessary. Rating scales 1–7 ranging from entire disagreement to entire agreement.

Table 2
Means/frequencies, standard deviations, and Pearson correlations of the study variables (N = 202).

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Female gender[a]																
2	Reasoning skills	−0.09															
3	Mathematical skills	**−0.16**	−0.02														
4	Language skills	**0.23**	−0.00	0.09													
5	Ability to program[a]	0.18	−0.06	−0.03	0.03												
6	Computer literacy	**−0.46**	**0.46**	**0.16**	−0.10	−0.17											
7	CT self-concept	**−0.42**	0.18	**0.28**	−0.09	−0.09	**0.32**										
8	Self-determined CT motivation	**−0.31**	**0.19**	**0.21**	**−0.19**	−0.10	0.26	**0.56**									
9	SECS	0.06	0.02	**−0.15**	**0.21**	−0.01	−0.02	−0.07	−0.16								
10	Migration background[a]	−0.06	0.07	−0.03	−0.10	−0.10	0.12	0.10	0.17	**−0.37**							
11	Formal learning in school year	−0.01	−0.10	0.06	−0.09	**−0.21**	−0.10	0.17	0.15	−0.12	0.13						
12	Past computer science instruction	−0.09	0.07	0.17	0.02	0.05	**0.24**	**0.24**	0.13	−0.02	−0.10	0.03					
13	Duration of computer use	**−0.25**	0.02	0.06	−0.10	**−0.17**	**0.31**	**0.35**	0.14	−0.05	0.08	0.09	0.06				
14	CT t1	**−0.41**	**0.37**	**0.28**	−0.01	0.00	**0.32**	**0.61**	**0.44**	0.05	0.05	**0.21**	−0.02	**0.26**			
15	CT t2	**−0.49**	**0.31**	**0.27**	0.01	−0.10	**0.44**	**0.58**	**0.38**	0.14	0.11	0.12	0.04	**0.36**	**0.68**		
16	CT t3	**−0.35**	**0.50**	**0.23**	0.02	−0.09	**0.52**	**0.49**	**0.31**	0.13	0.17	**0.15**	0.02	**0.37**	**0.68**	**0.65**	
M/frequency[a]		0.56	3.72	4.69	4.77	0.77	7.29	4.04	3.32	50[b]	0.46	2.63	2.89	3.16	50[b]	48.1	45.8
SD		0.50	1.99	0.75	0.54	0.41	4.53	1.43	1.52	10[b]	0.50	0.93	1.20	1.20	10[b]	11.0	11.4

Note. Correlations in bold are significant considering the cluster structure of the data: $p < .05$ (two-tailed).

[a] For dummy coded variables the percentage of category ‘1’ is reported: Gender (0 = male, 1 = female); Ability to program (0 = no, 1 = yes); Migration background (0 = both parents born in Switzerland, 1 = at least one parent not born in Switzerland).

[b] standardized to M = 50, SD = 10.

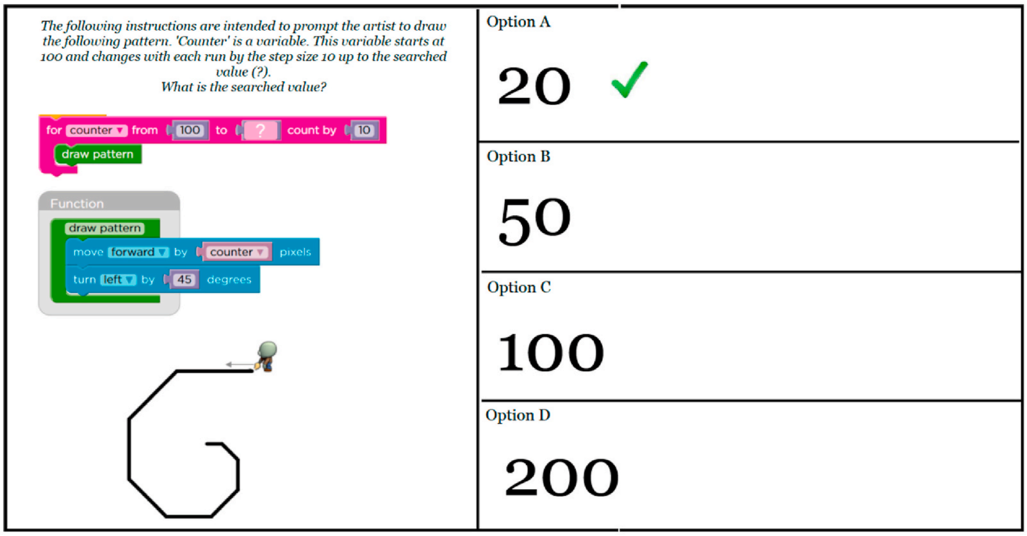


Fig. 1. Sample item of the CT performance test used in the study (Guggemos, Seufert, & Román-González, 2020).

In terms of gender, students' self-reports were utilized (Margulieux, Ketenci, & Decker, 2019). Reasoning skills were measured using the short version of the Hagen Matrices Test (HMT-S) (Heydasch, Haubrich, & Renner, 2017). It comprises six multiple-choice items and covers intelligence facet reasoning (Schneider & McGrew, 2012). Students' self-reported grades in mathematics and German (the general language spoken at school and the language of all measurement instruments) for the previous school year act as a proxy for mathematical and language skills. School grades in Switzerland range from 1 to 6, where 6 represents the best possible grade. In general, self-reported grades may have the same usability as actual grades (Gonyea, 2005). To capture ability to program, we asked students 'Are you able to program (e.g., Java or Python)?' (yes/no). Computer literacy was measured with the dimension 'practical computer knowledge' of the revised Computer Literacy Inventory (INCOBI-R) (Richter et al., 2010). It comprises 20 multiple-choice questions (performance test) that address situations that may occur when working with computers. A sample item is: 'You want to prevent other people from following your navigation behavior on the Internet. What measure contributes to this?'

CT self-concept was measured with six items taken from Köller, Daniels, Schnabel, and Baumert (2000) that we adapted to the context of our study. A sample item is: 'Generally, I can solve that kind of tasks well'. Self-determined motivation consists of the facets: identified, intrinsic, and interested (Prenzel, Drechsel, & Kramer, 1998). Each facet comprises three items. Sample items are (Prenzel, Drechsel, & Kramer, 1998): 'I get involved because I want to get a little closer to my own goals' (identified), 'Performing such tasks is fun' (intrinsic), and 'I am so fascinated that I fully committed myself' (interested). In line with Howard, Gagné, and Bureau (2017), we opted for capturing self-determined motivation with one factor.

To form the variable SECS, students' self-reports were used. From indications regarding parents' professions, e.g., 'mechanical engineer', we derived the 'International Socio-economic Index of Occupational Status' (ISEI) using the database of Ganzeboom and Treiman (2012). The ISEI ranges from 16 (e.g., cleaners in offices) to 90 (judges) and reflects income as well as educational level (Ganzeboom, Graaf, & Treiman, 1992). Furthermore, we used the number of books in the home on a six-point scale ranging from 0 to 10 to 500+ (an illustration was provided) as a proxy for the available cultural goods at home (OECD, 2017). By means of a principal component analysis, we extracted one factor. This approach is consistent with ICILS and PISA (Fraillon et al., 2019; OECD, 2017). However, we used the ISEI of both parents instead of the highest ISEI. This means that a family where both parents are judges is assigned a higher SECS than a family where the mother is a judge and the father is a plumber. If the highest ISEI was utilized to form SECS, both families would be assigned the same SECS. Moreover, we did not include parents' educational level as an individual variable; the ISEI also reflects the educational level (Ganzeboom et al., 1992). As for the CTt, we scaled the SECS to $M = 50$ and $SD = 10$. The dichotomous variable 'Migration background' equals zero (0) if students reported that both parents were born in Switzerland and one (1) if at least one parent was born outside Switzerland (Retelsdorf et al., 2011).

For capturing formal learning opportunities during the school year on a general level in t3, we drew on the 'generic skill' scale of Byrne and Flood (2003). A sample item is: 'The courses taken during the school year have promoted my ability to solve such problems.' The students in our sample attended 90 min of computer science instruction per week. Hence, it is not possible to capture formal learning opportunities during the school year by the number of computer science lessons due to a lack of variance. For past formal learning opportunities, we utilized the self-reported weekly number of past computer science lessons. Concerning informal learning opportunities, the self-reported computer use by students (PC or laptop) was used (OECD, 2017).

3.2.3. Reliability assessment and taken actions

For assessing the internal consistency reliability (reliability) of our measures, we used Revelle's Omega Total as it is superior to Cronbach's Alpha (McNeish, 2018). Since Cronbach's Alpha is widely used, however, we report it along with Omega. A reliability of greater than 0.7 may be sufficient (Hair, Risher, Sarstedt, & Ringle, 2019). Table 1 summarizes the operationalization of the CT predictors, reports their reliability, and shows what actions are taken.

Ability to program may be problematic because it is a self-assessment based on a single binary item. However, using a latent variable approach, unreliability can be modeled. Following Wang and Wang (2020, pp. 144–150), we restricted the error variance of the predictor Ability to program to Var(A) $(1 - \rho_A)$, where Var(A) is the observed variance of Ability to program and ρ_A its reliability. To determine the reliability, we drew from research about single item reliability. Postmes, Haslam, and Jans (2013) carried out a meta-analysis and reported a reliability (Cronbach's Alpha) of single item measures between 0.14 and 0.68. As a proxy for ρ_A, we took the midpoint of this range (0.41). Needless to say, this value can only act as a rough approximation. We will therefore come back to this point in the discussion and limitation section.

3.3. Data analysis

3.3.1. Missing data and outliers

The students who were present during the data collection fully answered the questions. In t1 and t2, absence of data can be attributed to potentially random events, mainly sick leave. In t3, an entire class (n = 23) dropped out due to scarcity of time. Since the dropout was not caused by CT-related reasons, e.g., poor performance in the preceding tests, we regard this data as missing at random. The final sample comprises N = 202, N = 198, and N = 177 students in t1, t2, and t3, respectively. Following the simulation study of Shin, Davison, and Long (2017), we relied on full information maximum likelihood estimation to handle the missing data. We also checked for outliers by means of Mahalanobis distances (Penny, 1996); no student was excluded.

3.3.2. Longitudinal data analysis

To answer the research question, we used latent growth curve modeling by means of the R-package lavaan 0.6–3 (Rosseel, 2012). Since our data is clustered (students nested in classes) and we are interested in the individual level, the basis for statistical inferences are cluster-robust standard errors (McNeish, Stapleton, & Silverman, 2017). In light of this, we report χ^2 statistics after considering the Satorra Bentler correction: SB-χ^2 (Satorra & Bentler, 2010). First, we created an intercept only model: the intercept as a latent variable is measured by the CTt scores of t1–t3 with all three loadings restricted to one corresponding to the three time points (August, January, and June). Second, we created an unconditional linear growth model by further inserting a slope as a latent variable. To this end, we constrained the loadings on the intercept to one; the loadings of the slope are constrained to 0, 1, and 2, also corresponding to t1–t3. By comparing the intercept only model with the linear growth model, we can gauge if there is significant change in CT over the school year. Third, we assessed if the assumption of a linear growth is justified. In order to do this, we restricted the loadings on the slope to 0 in t1 and 1 in t3; the loading for t2 is freely estimated. An estimated loading of greater (smaller) than 0.5 would imply that the expected growth from t1 to t2 is greater (smaller) than the growth from t2 to t3.

To examine the association of the predictors with the level (intercept) and growth (slope) of CT, we utilized a two-step strategy (Retelsdorf et al., 2011). First, each predictor was individually included in a simple conditional model to test its association with level and growth. Fig. 2 depicts the simple conditional model.

Second, we tested all predictors at once in a full conditional model to determine which ones uniquely contribute to explaining level, as well as growth, and to test the hypotheses.

4. Results

4.1. Descriptive results and correlations

Table 2 shows the Pearson correlations (r) of the considered variables as well as their mean or frequency and SD. The ISEI equals on average 57.34 (SD = 25.11) for the father and 51.60 (SD = 24.63) for the mother. In the median household, 101–200 books are available. The median student uses the computer 31–60 min per day. The largest correlation with CT appears for CT self-concept in t1 and t2 and for computer literacy in t3.

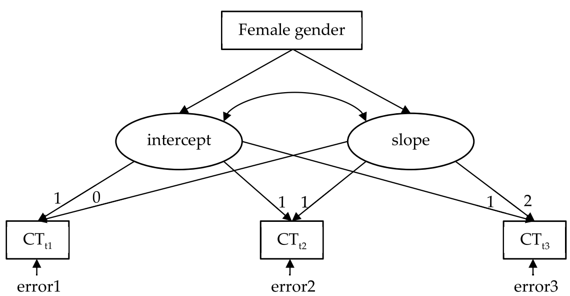


Fig. 2. Simple conditional model with intercept and slope. Female gender as an example for a time-invariant covariate.

4.2. Longitudinal data analysis

The intercept only model yielded an insufficient fit (Hu & Bentler, 1999): SB-$\chi^2(4) = 9.21$, $p = .056$, CFI = 0.900, TLI = 0.925, RMSEA = 0.185, SRMR = 0.084. The unconditional linear growth model yielded a significantly better fit: ΔSB-$\chi^2(3) = 8.85$, $p = .031$. A non-linear growth model did not yield a significant improvement in fit ($p = .941$); hence, we used a linear growth model for all our analyses. In this model, the slope is not significantly negative ($b = -1.70$, $p = .067$). Since the mean of CT in t1 (intercept) is standardized to 50 with a SD of 10, this equals a non-significant decrease of CT of about one-sixth of the standard deviation in t1. We will address this finding in the discussion section. The variance of the intercept equals 73.92, and the variance of the slope, 2.52. Intercept and slope are not significantly correlated ($r = 0.08$, $p = .896$).

The full conditional model showed a decent fit: SB-$\chi^2(27) = 29.44$, $p = .340$, CFI = 0.992, TLI = 0.984, RMSEA = 0.027, SRMR = 0.039. Table 3 presents both the simple and full conditional model. At the top, the prediction for the initial level can be found and, at the bottom, the prediction for growth.

Table 4 summarizes what hypotheses were accepted or rejected based on the full conditional model. It also indicates the cumulative explained variance grouped by the three areas of the conceptual framework.

5. Discussion

5.1. Findings

5.1.1. Overall findings

The students in the sample had 90 min of computer science instruction per week during the school year of the investigation. Surprisingly, we observed a non-significant negative growth of CT. Cohen's d equals 0.16 from t1 to t2 and 0.20 from t2 to t3, which would be a marginal and a small effect, respectively. The reason for this finding may be a decrease in test motivation during the school year. As a robustness check, we formed a proxy for lack of test motivation by means of two items measuring the construct amotivation on a seven-point scale of rating, captured at all three time points (Prenzel, Drechsel, & Kramer, 1998): 'When performing the tasks, I disengaged' and 'My mind was elsewhere when I was performing the tasks.' Cronbach's Alpha is greater than 0.74 at all three time points. We observed an increase in lack of test motivation from t1 ($M = 3.36$, $SD = 1.66$) to t2 ($M = 4.18$, $SD = 1.47$) to t3 ($M = 4.75$,

Table 3

Conditional models for CT performance (N = 202).

H	Predictor	Simple conditional model				Full conditional model			
		Est.	Std. Est.	s.e.[a]	p-value	Est.	Std. Est.	s.e.[a]	p-value
					Predicting initial level				
1	Female gender	**−10.12**	**−0.60**	**1.41**	**0.000**	**−4.32**	**−0.24**	**1.54**	**0.005**
2	Reasoning skills	**2.13**	**0.47**	**0.52**	**0.000**	**1.40**	**0.30**	**0.32**	**0.000**
3	Mathematical skills	**3.87**	**0.33**	**0.74**	**0.000**	**1.57**	**0.13**	**0.48**	**0.001**
4	Language skills	−1.32	−0.08	1.79	0.463	1.01	0.06	1.50	0.499
5	Ability to program	0.03	0.00	2.49	0.990	**5.58**	**0.16**	**2.47**	**0.024**
6	Computer literacy	**0.84**	**0.44**	**0.19**	**0.000**	−0.16	−0.08	0.16	0.325
7	CT self-concept	**4.14**	**0.71**	**0.24**	**0.000**	**2.10**	**0.34**	**0.37**	**0.000**
8	Self-determined CT motivation	**2.89**	**0.53**	**0.45**	**0.000**	**0.93**	**0.16**	**0.39**	**0.017**
9	Parents' SECS	0.05	0.05	0.08	0.569	**0.14**	**0.16**	**0.06**	**0.019**
10	Migration background	2.41	0.14	1.39	0.083	**1.93**	**0.11**	**0.54**	**0.000**
11	Formal learning in school year	−0.20	−0.03	0.77	0.797	−0.45	−0.06	0.62	0.472
12	Past computer science instruction	**2.43**	**0.27**	**0.86**	**0.005**	**0.99**	**0.11**	**0.45**	**0.026**
13	Duration of computer use	**2.09**	**0.35**	**0.45**	**0.000**	**0.81**	**0.13**	**0.26**	**0.002**
					Predicting growth				
1	Female gender	0.57	0.18	1.17	0.626	1.74	0.34	1.14	0.125
2	Reasoning skills	0.33	0.37	0.19	0.092	0.24	0.18	0.13	0.074
3	Mathematical skills	−0.13	−0.06	0.56	0.820	−0.13	−0.04	0.41	0.752
4	Language skills	0.72	0.23	0.68	0.291	0.62	0.13	0.51	0.224
5	Ability to program	−1.37	−0.34	0.71	0.053	−1.17	−0.12	2.04	0.567
6	Computer literacy	**0.25**	**0.58**	**0.10**	**0.011**	**0.33**	**0.58**	**0.10**	**0.001**
7	CT self-concept	−0.10	−0.09	0.28	0.719	−0.11	−0.06	0.30	0.727
8	Self-determined CT motivation	−0.30	−0.28	0.29	0.297	−0.40	−0.25	0.24	0.091
9	Parents' SECS	0.06	0.33	0.04	0.207	0.06	0.24	0.04	0.165
10	Migration background	0.81	0.25	0.65	0.218	0.81	0.16	0.49	0.096
11	Formal learning in school year	0.23	0.18	0.21	0.279	**0.49**	**0.23**	**0.13**	**0.000**
12	Past computer science instruction	−0.38	−0.22	0.44	0.379	−0.58	−0.21	0.47	0.217
13	Duration of computer use	**0.51**	**0.42**	**0.18**	**0.004**	**0.46**	**0.26**	**0.19**	**0.014**

Note. Std. Est. = Standardized Estimate.

[a] Cluster robust standard errors at the class level. Estimates in bold are significant (two-tailed) at the 5% level. All variance inflation factors (VIF) <1.91.

Table 4
Summary of the tested hypotheses and explained variance (N = 202).

Area	H	Predictors (expected sign)	Level		Growth	
			Accept	R^2	Accept	R^2
Student	1	Female gender (−)	✓**	68.7%	X	55.0%
	2	Reasoning skills (+)	✓***		X	
	3	Mathematic skills (+)	✓***		X	
	4	Language skills (+)	X		X	
	5	Ability to program (+)	✓*		X	
	6	Computer Literacy (+)	X		✓**	
	7	Self-concept (+)	✓***		X	
	8	Self-determined motivation (+)	✓*		X	
Home environment	9	Parents' SECS (+)	✓*	70.3%	X	56.8%
	10	Migration background (−)	X		X	
Learning opportunities	11	Current learning opportunities (+)	n/a	70.4%	✓***	61.2%
	12	Past computer science instruction (+)	✓*		X	
	13	Duration of computer use (+)	✓**		✓*	

Note. ✓ = accepted, X = rejected. *** = significant at the 0.1% level (two-tailed), ** = significant at the 1% level (two-tailed), and * = significant at the 5% level (two-tailed). N/a means that no hypothesis was developed. R^2 indicates the cumulative explained variance.

SD = 1.42), equivalent to a Cohen's d of 0.52 and 0.39, respectively. Starting from a non-negative slope in the unconditional growth model (b = −1.70, p = .067), inserting amotivation as a time-varying covariate leads to a non-significant positive slope (b = 2.17, p = .108). However, considering amotivation as a time-varying covariate in the full conditional model does not change the interpretation (sign and significance) of any of the reported estimates. This may indicate the robustness of our findings concerning decreasing test motivation. Another point that could contribute to explaining the finding is students' self-reports about experiencing rather low learning opportunities during the school year (M = 2.63, SD = 0.93 on a seven-point scale). Computer science was newly introduced as a curriculum subject in the school year of our investigation. Implementing a new curriculum is a complex endeavour that takes time (Altrichter, 2005). Hence, the ideas of the curriculum had potentially not yet been fully implemented at that time. Moreover, a one-year period of time may not be long enough to observe substantial change; learning might not be a linear process for all students.

5.1.2. Antecedents – student

Females show CT scores that are about one standard deviation lower than those of males (simple conditional model). Since there is evidence for the fairness of CTt (Guggemos, Seufert, & Román-González, 2019), this finding may not be the result of a lack of test fairness. Of the gender difference, more than 50% can be attributed to females' lower CT self-concept, lower computer literacy, and lower self-determined motivation (simple vs. full conditional model). However, even when considering all covariates in the full conditional model, a negative association of 'female gender' and CT level remains. Román-González et al. (2017) argued that the CT gender gap may increase with age. This, indeed, may be consistent with the available evidence: the ICILS reported only minor gender differences for 8th graders; our sample of 11th graders revealed a substantial gap. The question that remains is what variables, like personality traits, might explain these findings? We agree with Shute et al. (2017) that gender differences deserve further research.

Language skills do not show a positive association with the CT level in the simple conditional model or in the full conditional model. This finding is contrary to Román-González et al. (2018). Since the study of those authors also rely on the CTt for measuring CT, one reason could be that the students in our sample show a sufficient level of language skills to perform the CTt items and, above this level, language skills no longer have a positive influence on the CTt performance. To support this line of argumentation, we split up our sample in a low and a high language skill group using the median (= 5). Since we did not find a (stronger) positive association among the students in the low language skill group in comparison to the high language skill group, all students in our sample may have sufficient language skills to perform the CTt items. On a more general level, however, further research may be necessary to obtain a better understanding about how exactly language skills positively influence CT. Regarding the ability to program, the association with CT level is only significant in the full conditional model and if measurement error is considered (b = 5.58, p = .024). If measurement error is not taken into account, the association is not significant (b = 2.13, p = .068). As a robustness check, we carried out a sensitivity analysis. It indicated that up to an unrealistically high reliability of 0.7, the association remains significant at the 5% level. Overall, however, the relationship between text-based and block-based programming, like in our CT test, may not be as straightforward as often assumed. Programming skills required in text-based languages are only partially overlapping those required in block-based languages (Mladenović, Boljat, & Žanko, 2018; Weintrop & Wilensky, 2017). The association between computer literacy and level of CT is not statistically significant in the full conditional model, but in the simple conditional model. The main reasons are the inclusion of CT self-concept and reasoning skills in the full conditional model; both are positively correlated with computer literacy.

With regard to motivation, both components, self-concept and self-determined motivation, are significant predictors of the CT level. CT self-concept has the strongest association with the level of CT in both the simple and full conditional model. In sum, these findings may indicate the usefulness of the EVM for capturing motivation in CT research. Considering motivation might be especially important for the understanding of gender differences. Simple comparisons may overestimate the effect of gender because a substantial part is explained by female students' lower CT self-concept and self-determined motivation.

From the student antecedents, only computer literacy is a significant predictor for growth. Since computer literacy does not uniquely predict the CT level, specific circumstances during the school year could be the reason: CT may be instructed in a way that allows students with higher levels of computer literacy to benefit more from instruction. In fact, students worked with computers or digital devices almost all time during the computer science lessons. A multigroup analysis that compares students with low (0–6 correct answers in the computer literacy test), medium (7–12 correct answers), and high (13–20 correct answers) computer literacy revealed for the high computer literacy group a significant positive slope: $b = 1.92$, $p = .021$. The slopes of the medium and low group are not significantly different ($p = .618$). For the pooled medium/low group, the slope equals $b = -2.14$ ($p = .014$). In both groups the estimated slopes are higher when amotivation is considered as a time-varying covariate: $b = 6.45$, $p < .001$ (high computer literacy group) and $b = 1.23$, $p = .385$ (medium/low computer literacy group).

5.1.3. Antecedents – home environment

Home environment comprises SECS and migration background. These variables are (in line with large-scale studies like ICILS and PISA) negatively correlated ($r = -0.37$, $p < .001$). Both predictors do not show a significant association with the CT level in the simple conditional model, but do so in the full conditional model. Concerning SECS, this can mainly be attributed to its negative correlation with self-determined motivation. For migration background, there is no specific reason. The finding for a positive association of SECS with the level of CT is in line with the ICILS. Other than hypothesized, however, a migration background has a positive association with the CT level in our study. The explanation might be that general reasons for migrant students' lower performance, like language problems, are not pertinent in the context of our study. Our sample comprises students from the 'Gymnasium Helveticum' where students are selected by means of a rigorous admission test covering, among others, language and mathematical skills. In terms of reading, the 2018 PISA study showed, for Switzerland, that migrants do not perform worse if control variables like SECS are considered (Konsortium PISA.ch, 2018). In our case, students with a migration background might perform even better than students without such a background because these students, in our specific circumstances, may be very competitive and from families with a high affinity towards education. Overall, consistent with the ICILS, home environment seems to be an important variable when investigating CT. As a robustness check, we also used a different measure for SECS. Instead of considering both parents ISEI, we used the highest ISEI of father and mother along with the number of books to form SECS. With one exception, the results are identical in terms of sign and significance at the 5% level: SECS is significantly positively associated with CT growth ($\beta = 0.33$, $p = .035$). In the specification reported in Table 3, home environment variables showed no significant association with CT growth.

5.1.4. Processes – learning opportunities

Learning opportunities in informal and formal settings have a positive association with growth of CT. The finding of a positive association between duration of computer use and CT might confirm the results of Werner et al. (2012) and complement the work of Durak and Saritepeci (2018), which did not find a positive influence from the general use of ICT. Overall, CT learning seems to be related to computing because duration of computer use and computer literacy are both positively associated with CT growth. Those two predictors are positively correlated ($r = 0.24$, $p < .001$). The perceived formal learning opportunities during the school year do not predict CT growth in the simple conditional model, but they do in the full conditional model. The main reason for this increase in the coefficient is the inclusion of computer literacy, which is negatively correlated with perceived learning opportunities during the school year.

5.2. Limitations

Our study is not without limitations. We discuss them following Cachero, Barra, Melia, and Lopez (2020) by addressing internal, external, construct, and conclusion validity.

Compared to randomized field experiments as the gold standard, internal validity is suboptimal. The reported associations cannot be interpreted as causal because we cannot rule out that variables not considered in our conceptual framework cause the associations. We tried to reduce this risk by a comprehensive review on variables that influence CT from a theoretical point of view. Another threat to the internal validity could be the drop out of one school class in t3. This drop out, however, was not due to the results in the preceding tests; the data may be missing at random.

A threat to external validity could be our sample. It is narrow in scope as it comprises only students from German-speaking Switzerland, and from one type of school. This sample is not representative for high-school students in general. For making general claims about those students, replication studies would be necessary. Moreover, computer science was newly introduced as a curriculum subject in the school year of investigation. Since implementing a curriculum is a complex and time consuming endeavour (Altrichter, 2005), the results on CT growth could be different if the study was replicated in the future. Moreover, it is doubtful to what extent our findings on CT can be generalized to text-based programming due to an only partial overlap with block-based programming languages.

In terms of construct validity, CT as the dependent variable is measured using the CTt, a unidimensional performance test that is by nature narrow in scope. Concerning Brennan and Resnick's (2012) framework, the CTt focuses mainly on computational concepts, partly on computational practices, and neglects computational perspectives. It is unlikely that the complex nature of CT can be captured by using a single instrument. Rather, a system of assessment tools may be necessary in order to gain a comprehensive picture of students' CT (Román-González et al., 2019; Tang, Yin, Lin, Hadad, & Zhai, 2020). In terms of the predictors, Ability to program was captured with only one question: 'Are you able to program (e.g., Java or Python)?' This question may be hard to answer because there is no clear-cut definition of ability to program. Moreover, a dichotomous item may be suboptimal because programming ability is probably a continuum. Using a more elaborate self-assessment instrument or a performance test in future research would be beneficial.

To tackle the inherent unreliability of our measure, we can only rely on a very rough approximation for its measurement error. However, up to an unrealistically high reliability of 0.7 of the binary single item, the association with the CT level remains significant at the 5% level. Besides this, self-reported grades could suffer from a self-serving bias (Campbell & Sedikides, 1999) and students may not be able or willing to fairly assess themselves.

Conclusion validity may be impaired as we focused on the individual level. Learning (of CT), however, is embedded in classrooms, schools, and an educational system (Fraillon et al., 2019). These higher-level variables may moderate relationships at the individual level. In particular the consideration of the class and the school level by means of a multi-level analysis could provide further insight into the level of CT and its growth. Nevertheless, our statistical inferences may be valid because we relied on cluster-robust standard errors (McNeish et al., 2017).

5.3. Implications

Overall, computer science instruction may be conducive to CT because there is a positive association with the number of attended computer science lessons in the past and CT. However, although the students in the sample attended 90 min of computer science instruction per week, they reported low levels of learning opportunities ($M = 2.63$, $SD = 0.93$ on a seven-point scale). Moreover, if decreasing test motivation is not considered, the level of CT growth over the school year was actually (non-significantly) negative. As Cachero et al. (2020) claim, CT may not develop naturally. Rather, a deliberate effort may be necessary for effectively fostering CT (within computer science instruction).

Motivation plays a crucial role in predicting the CT level. Hence, when selecting a method for CT instruction, motivational aspects may be considered. In this vein, Repenning (2017) argues that professional programming languages may demotivate students and it might be preferable to rely on visual programming languages in CT instruction. Professional (text-based) programming languages might be regarded as too difficult (lack of CT self-concept) and not attractive enough (lack of self-determined motivation). As our research might indicate, both components of the EVM should be considered when selecting CT instructional means and carrying out research that compares text-based and block-based programming (Mladenović et al., 2018; Weintrop & Wilensky, 2017).

Initiatives that involve females in CT may be necessary because the gender gap does not seem to decrease with age. Rather, the opposite may be the case. Since CT is an important 21st-century skill prevalent in all areas of study and in the workplace (Yadav et al., 2018), initiatives to tackle this issue may be required. In terms of instruction, game-based approaches could be promising for engaging female interest, as well as students from minority groups (Repenning et al., 2015). Efforts involving female students in CT activities may start at the elementary level (e.g., Luo, Antonenko, & Davis, 2020).

6. Conclusion and outlook

Our study has provided new insight into predictors of CT because we used a sample of high-school students, an under-researched area. Using a longitudinal design allowed us to investigate predictors of CT level and growth in a naturalistic classroom setting. An explained variance of 70.4% for CT level and 61.2% for CT growth might indicate a good trade-off between the comprehensiveness and parsimony of our framework. Further (qualitative) research could aim at identifying more, not as yet considered, predictors of CT.

In terms of predictors, we showed that gender is important for explaining the level of CT within a high-school context. Language skills seem to have no influence either on the level or growth of CT. Since mathematical skills can predict CT level, CT may be more closely related with mathematical than with language skills. Concerning the ability to program, we find a significant association with CT if measurement error is considered. We suggest using a performance test or an elaborate self-assessment instrument for measuring programming ability in future research. Motivation, in the form of CT self-concept and self-determined motivation, plays an important role in explaining CT level and gender differences. The duration of computer (PC and laptop) use has a positive association with CT level and growth. Further research could investigate what kind of activities carried out during computer use are conducive to CT. Moreover, we showed that perceived learning opportunities during the school year under investigation are positively associated with CT growth. Here, follow-up questions should include the kind of activities and courses responsible for such a perception.

Acknowledgement

The author has received a postdoctoral fellowship for carrying out this research from the Basic Research Fund at the University of St.Gallen (no. 1031542). Our thanks go to the coordinators and administrators at the three schools that participated in this study. We also thank Marcos Román-González who shared with us the Computational Thinking Test (CTt) and was involved in the process of adapting the CTt for this study. Finally, we are very much indebted to Frank Fischer, Michael Sailer, and Sabine Seufert for their help and insightful suggestions during the conduct of this research. Readers who are interested in replicating the study may contact the author.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2020.104060.

References

AERA, A. P. A., & NCME. (2014). *Standards for educational and psychological testing*. Washington D.C.: American Educational Research Association.

Altrichter, H. (2005). Curriculum implementation – limiting and facilitating factors. In P. Nentwig, & D. Waddington (Eds.), *Making it relevant: Context based learning of science* (pp. 35–62). Münster: Waxmann.

Ambrosio, A. P., Xavier, C., & Georges, F. (2015). Digital ink for cognitive assessment of computational thinking. *IEE Frontiers in Education Conference*, 1520–1526. https://doi.org/10.1109/FIE.2014.7044237

Angeli, C., Voogt, J., Fluck, A., Webb, M., Cox, M., Malyn-Smith, J., et al. (2016). A K-6 computational thinking curriculum framework: Implications for teacher knowledge. *Journal of Educational Technology & Society, 19*(3), 47–57. https://eric.ed.gov/?id=EJ1107187.

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems, 75*, 661–670. https://doi.org/10.1016/j.robot.2015.10.008

Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning and Leading with Technology, 38*(6), 20–23. https://eric.ed.gov/?id=EJ918910.

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12. *ACM Inroads, 2*(1), 48–54. https://doi.org/10.1145/1929887.1929905

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking*. Vancouver, BC: American Educational Research Association Meeting. Canada, 1–25 https://www.media.mit.edu/publications/new-frameworks-for-studying-and-assessing-the-development-of-computational-thinking/.

Buitrago Flórez, F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., & Danies, G. (2017). Changing a generation's way of thinking: Teaching computational thinking through programming. *Review of Educational Research, 87*(4), 834–860. https://doi.org/10.3102/0034654317710096

Byrne, M., & Flood, B. (2003). Assessing the teaching quality of accounting programmes: An evaluation of the course experience questionnaire. *Assessment & Evaluation in Higher Education, 28*(2), 135–145. https://doi.org/10.1080/02602930301668

Cachero, C., Barra, P., Melia, S., & Lopez, O. (2020). Impact of programming exposure on the development of computational thinking capabilities: An empirical study. *IEEE Access, 8*, 72316–72325. https://doi.org/10.1109/ACCESS.2020.2987254

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology, 3*(1), 23–43. https://doi.org/10.1037/1089-2680.3.1.23

Cutumisu, M., Adams, C., & Lu, C. (2019). A scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology, 28*(6), 651–676. https://doi.org/10.1007/s10956-019-09799-3

Del Olmo-Muñoz, J., Cózar-Gutiérrez, R., & González-Calero, J. A. (2020). Computational thinking through unplugged activities in early years of primary education. *Computers & Education, 150*. https://doi.org/10.1016/j.compedu.2020.103832

Duncan, C., & Bell, T. (2015). A pilot computer science and programming course for primary school students. In A. Mühling, A. Ruf, & P. Hubwieser (Eds.), *Design and first results of a psychometric test for measuring basic programming abilities* (pp. 39–48). New York, NY: ACM. https://doi.org/10.1145/2818314.2818328.

Durak, H. Y., & Saritepeci, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education, 116*, 191–202. https://doi.org/10.1016/j.compedu.2017.09.004

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA) https://www.iea.nl/publications/study-reports/preparing-life-digital-world.

Ganzeboom, H. B. G., Graaf, P. M. de, & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Ganzeboom, H. B. G., & Treiman, D. J. (2012). *International stratification and mobility file: Conversion tools*. http://www.harryganzeboom.nl/isco08/index.htm.

Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. *New Directions for Institutional Research*, (127), 73–89. https://doi.org/10.1002/ir.156

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Grover, S., Pea, R., & Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education, 25*(2), 199–237. https://doi.org/10.1080/08993408.2015.1033142

Grover, S., Pea, R., & Cooper, S. (2016). Factors influencing computer science learning in middle school. In C. Alphonce, J. Tims, M. Caspersen, & S. Edwards (Eds.), *Proceedings of the 47th ACM technical symposium on computing science education* (pp. 552–557). New York, NY: ACM. https://doi.org/10.1145/2839509.2844564.

Guggemos, J., Seufert, S., & Román-González, M. (2019). Measuring computational thinking - Adapting a performance test and a self-assessment instrument for German-speaking countries. In *Proceedings of the 16th International Conference Cognition and Exploratory Learning in the Digital Age (CELDA)* (pp. 183–191). https://doi.org/10.33965/celda2019_201911L023

Guggemos, J., Seufert, S., & Román-González, M. (2020). Validity evidence for the "Computational Thinking Test" at the upper secondary level using item response theory and confirmatory factor analysis. In *AERA Annual Meeting San Francisco, CA*. http://tinyurl.com/wdpncar.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Heydasch, T., Haubrich, J., & Renner, K.-H. (2017). The short version of the Hagen Matrices test (HMT-S): 6-Item induction intelligence test. *Methoden, Daten, Analysen, 7*(2), 183–202. https://doi.org/10.12758/MDA.2013.011

Howard, J. L., Gagné, M., & Bureau, J. S. (2017). Testing a continuum structure of self-determined motivation: A meta-analysis. *Psychological Bulletin, 143*(12), 1346–1377. https://doi.org/10.1037/bul0000125

Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296–310. https://doi.org/10.1016/j.compedu.2018.07.004

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Ilic, U., Haseski, H. I., & Tugtekin, U. (2018). Publication trends over 10 years of computational thinking research. *Contemporary Educational Technology, 9*(2), 131–153. https://doi.org/10.30935/cet.414798

Israel, M., Pearson, J. N., Tapia, T., Wherfel, Q. M., & Reese, G. (2015). Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis. *Computers & Education, 82*, 263–279. https://doi.org/10.1016/j.compedu.2014.11.022

Ketenci, T., Calandra, B., Margulieux, L., & Cohen, J. (2019). The relationship between learner characteristics and student outcomes in a middle school computing course: An exploratory analysis using structural equation modeling. *Journal of Research on Technology in Education, 51*(1), 63–76. https://doi.org/10.1080/15391523.2018.1553024

Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Course selections of girls and boys in mathematics: The role of academic self-concept and interest. *Zeitschrift für Pädagogische Psychologie, 14*(1), 26–37. https://doi.org/10.1024//1010-0652.14.1.26

Kong, S.-C., Chiu, M. M., & Lai, M. (2018). A study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education. *Computers & Education, 127*, 178–189. https://doi.org/10.1016/j.compedu.2018.08.026

Konsortium, P. I. S. A. (2018). *Schülerinnen und Schüler der Schweiz im internationalen Vergleich [Students in Switzerland in international comparison]*. Bern, Genf: Konsortium PISA.ch https://www.sbfi.admin.ch/dam/sbfi/de/dokumente/2019/12/bericht-pisa-2018.pdf.download.pdf/pisa-2018-bericht_d.pdf.

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior, 72*, 558–569. https://doi.org/10.1016/j.chb.2017.01.005

Leighton, J. P. (2009). *The nature of reasoning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511818714

Luo, F., Antonenko, P. D., & Davis, E. C. (2020). Exploring the evolution of two girls' conceptions and practices in computational thinking in science. *Computers & Education, 146*. https://doi.org/10.1016/j.compedu.2019.103759

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior, 41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Margulieux, L., Ketenci, T. A., & Decker, A. (2019). Review of measurements used in computing education research and suggestions for increasing standardization. *Computer Science Education, 29*(1), 49–78. https://doi.org/10.1080/08993408.2018.1562145

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology, 108*(3), 424–437. https://doi.org/10.1037/edu0000061

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140. https://doi.org/10.1037/met0000078

Mladenović, M., Boljat, I., & Žanko, Ž. (2018). Comparing loops misconceptions in block-based and text-based programming languages at the K-12 level. *Education and Information Technologies, 23*(4), 1483–1500. https://doi.org/10.1007/s10639-017-9673-3

Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P., Jr. (2007). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research (167-188)*. Thousand Oaks: Sage.

OECD. (2015). *Immigrant students at school: Easing the journey towards integration*. Paris: OECD. https://doi.org/10.1787/9789264249509-en

OECD. (2017). *PISA 2015 ttechnical report*. Paris: OECD Publishing. http://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf.

Pakpahan, E., Hoffmann, R., & Kröger, H. (2017). Statistical methods for causal analysis in life course research: An illustration of a cross-lagged structural equation model, a latent growth model, and an autoregressive latent trajectories model. *International Journal of Social Research Methodology, 20*(1), 1–19. https://doi.org/10.1080/13645579.2015.1091641

Penny, K. I. (1996). Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance. *Applied Statistics, 45*(1), 73–81. https://doi.org/10.2307/2986224

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology, 52*(4), 597–617. https://doi.org/10.1111/bjso.12006

Prenzel, M., Drechsel, B., & Kramer, K. (1998). Lernmotivation im kaufmännischen unterricht: Die sicht von Auszubildenden und lehrkräften [Learning motivation in business education: The perspective of trainees and teachers]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik, special issue, 14*, 169–187.

Repenning, A. (2017). Moving beyond syntax: Lessons from 20 years of blocks programing in AgentSheets. *Journal of Visual Languages and Sentient Systems, 3*(1), 68–91. https://doi.org/10.18293/VLSS2017-010

Repenning, A., Webb, D. C., Koh, K. H., Nickerson, H., Miller, S. B., Brand, C., … Repenning, N. (2015). Scalable game design: A strategy to bring systemic computer science education to schools through game design and simulation creation. *ACM Transactions on Computing Education, 15*(2), 1–31. https://doi.org/10.1145/2700517

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction, 21*(4), 550–559. https://doi.org/10.1016/j.learninstruc.2010.11.001

Richter, T., Naumann, J., & Horz, H. (2010). A revised version of the computer literacy inventory (INCOBI-R). *Zeitschrift für Padagogische Psychologie, 24*(1), 23–37. https://doi.org/10.1024/1010-0652/a000002

Román-González, M. (2015). *Computational thinking test: Design guidlines and content validation*. https://doi.org/10.13140/RG.2.1.4203.4329

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In S.-C. Kong, & H. Abelson (Eds.), *Computational thinking education* (pp. 79–98). Singapore: Springer. https://doi.org/10.1007/978-981-13-6528-7_6.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior, 72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Extending the nomological network of computational thinking with non-cognitive factors. *Computers in Human Behavior, 80*, 441–459. https://doi.org/10.1016/j.chb.2017.09.030

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–35. https://doi.org/10.18637/jss.v048.i02

Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education, 8*(3), 259–278. https://doi.org/10.2304/rcie.2013.8.3.259

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology, 111*(5), 764–792. https://doi.org/10.1037/edu0000314

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (3rd ed., pp. 99–144). New York: Guilford Publications.

Seidel, T., Rimmele, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction, 15*(6), 539–556. https://doi.org/10.1016/j.learninstruc.2005.08.004

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. https://doi.org/10.2307/1170010

Shin, T., Davison, M. L., & Long, J. D. (2017). Maximum likelihood versus multiple imputation for missing data in small longitudinal samples with nonnormality. *Psychological Methods, 22*(3), 426–449. https://doi.org/10.1037/met0000094

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Sneider, C., Stephenson, C., Schafer, B., & Flick, L. (2014). Computational thinking in high school science classrooms. *The Science Teacher, 81*(5), 10–15. https://doi.org/10.2505/4/tst14_081_05_53

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education, 148*. https://doi.org/10.1016/j.compedu.2019.103798

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies, 20*(4), 715–728. https://doi.org/10.1007/s10639-015-9412-6

Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). Hoboken, NJ: Wiley.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627

Webb, M. E., Bell, T., Davis, N., Katz, Y. J., Fluck, A., Sysło, M. M., … Brodnik, A. (2018). Tensions in specifying computing curricula for K-12: Towards a principled approach for objectives. *IT - Information Technology, 60*(2), 59–68. https://doi.org/10.1515/itit-2017-0017

Weintrop, D., & Wilensky, U. (2017). Comparing block-based and text-based programming in high school computer science classrooms. *ACM Transactions on Computing Education, 18*(1), 1–25. https://doi.org/10.1145/3089799

Werner, L., Denner, J., & Campe, S. (2012). *The fairy performance assessment: Measuring computational thinking in middle school* (pp. 215–220). ACM Technical Symposium on Computer Science Education. https://doi.org/10.1145/2157136.2157200

Wigfield, A., & Eccles, J. (2000). Expectancy-Value Theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 366*, 3717–3725. https://doi.org/10.1098/rsta.2008.0118, 1881.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving — more than reasoning? *Intelligence, 40*(1), 1–14. https://doi.org/10.1016/j.intell.2011.11.003

Yadav, A., Good, J., Voogt, J., & Fisser, P. (2018). Computational thinking as an emerging competence domain. In M. Mulder (Ed.), *Technical and vocational education and training. Competence-based vocational and professional education: Bridging the worlds of work and education* (pp. 1051–1067). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41713-4_49.

Yeh, K.-C., Xie, Y., & Ke, F. (2012). Teaching computational thinking to non-computing majors using spreadsheet functions. In *Frontiers in Education Conference (FIE)*. https://doi.org/10.1109/FIE.2011.6142980, 2011 (F3J-1-F3J-5).

Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education, 141*. https://doi.org/10.1016/j.compedu.2019.103607

Zhao, W., & Shute, V. J. (2019). Can playing a video game foster computational thinking skills? *Computers & Education, 141*. https://doi.org/10.1016/j.compedu.2019.103633

Studie 3:

Moser, L., Guggemos, J. & Seufert, S. (2021). Improving a MOOC to foster information literacy by means of a conjecture map. *International Journal of Learning Technology*, 16(1), 65–86.

Improving a MOOC to foster information literacy by means of a conjecture map

Luca Moser*, Josef Guggemos and
Sabine Seufert

Institute of Business Education and Educational Management,
Digital Learning and Corporate Learning,
University of St. Gallen,
9000 St. Gallen, CH, Switzerland
Email: luca.moser@unisg.ch
Email: josef.guggemos@unisg.ch
Email: sabine.seufert@unisg.ch
*Corresponding author

Abstract: Conjecture mapping allows for a combined study of learning and teaching and its interdependence in a given learning design. This contribution illustrates how we developed and improved a massive open online course (MOOC) to foster information literacy (IL) with upper secondary students in Switzerland, by means of conjecture mapping. Teachers often feel insecure when handling digital media pedagogically and therefore need tools to support them. Hence, the present educational design research focused on an easy-to-use digital learning design to foster IL. Interviews with 16 involved teachers indicated that the efficiency of the course design is a main success factor that should be reflected in the conjecture map. Furthermore, teachers who supported the learning process intensively and enhanced the MOOC with further tests and discussions were more successful, so we adapt the MOOC accordingly to provide even more support for the teachers.

Keywords: conjecture map; information literacy; massive open online course; MOOC; open education; secondary education; educational design research; learning technology; upper secondary education; digital competences.

Reference to this paper should be made as follows: Moser, L., Guggemos, J. and Seufert, S. (2021) 'Improving a MOOC to foster information literacy by means of a conjecture map', *Int. J. Learning Technology*, Vol. 16, No. 1, pp.65–86.

Biographical notes: Luca Moser is a doctoral student and Assistant Researcher in Business Education at the Institute for Educational Management and Educational Technologies at the University of St.Gallen (IBB-HSG), Switzerland. His research focuses on the design and implementation of digital learning designs, such as MOOCs and augmented reality learning settings.

Josef Guggemos is an Assistant Professor of Educational Technologies and Computational Thinking at the Institute of Educational Management and Technologies at the University of St. Gallen, Switzerland. He earned his Master's degree in Business Research (MBR) and PhD from the Munich School of Management, LMU Munich, Germany. His research interests are: computational thinking, social robots in education, competencies of teachers and human resource developers in the context of digital transformation, information literacy, and business education science.

Sabine Seufert is a Professor in Business Education and Head of the Institute for Educational Management and Educational Technologies at the University of St. Gallen (IBB-HSG), Switzerland. Her research covers the management of educational innovation such as artificial intelligence in education and the digital transformation in general. Other foci lie on the development of digital competences and technology-based learning.

This paper is a revised and expanded version of a paper entitled 'Developing a MOOC to foster information literacy (IL) by means of a conjecture map' presented at LTEC, Zamora, Spain, 15–18 July 2019.

1 Introduction

The competent handling of information [information literacy (IL)] is a key competence necessary for participating in an increasingly interconnected knowledge economy and society (Horton, 2013). IL might also be a precondition for lifelong learning (Gapski and Tekster, 2009).

The current generation of students are familiar with digital devices and have been using them their entire life (Prensky, 2011). However, the use of digital devices does not automatically enhance digital competences (Fraillon et al., 2019). These findings of the ICILS report 2018 confirms the results of two research projects we conducted in 2015. Swiss students on the upper secondary level revealed deficits on IL. Several international studies confirm these findings for students in other countries (Ng, 2012; Šorgo et al., 2017; Coklar et al., 2017).

Furthermore, our 2015 studies indicated that teachers systematically overestimate their students' IL. Against this background and considering the high importance of IL, we think it is crucial to foster upper secondary students' IL on a large scale and to raise teachers' awareness about students' IL deficits. Moreover, there is a lack of research that focuses on IL on the upper secondary level (Coklar et al., 2017; Koltay et al., 2016).

IL literature recommends that IL should be fostered at school (Argelagós and Pifarré, 2012; Koltay et al., 2016), but it has not been sufficiently incorporated in instruction (Jones-Kavalier and Flannigan, 2006). Reasons may be a lack of theoretically founded and empirically validated IL models, which are necessary for effectively fostering IL (Hartig and Jude, 2007), and a shortage of adequate IL learning material. Teachers need support and clarity on how to evaluate IL in their subjects (Lanning and Mallek, 2017). Moreover, the ability of teachers to use digital media in a classroom is of the utmost importance to teach IL. Many teachers, however, still lack a sufficient level of competences to design digital learning settings on their own (Fraillon et al., 2019) that combine the new requirements of digital, open, learner-centred settings with the fixed timetables.

In order to support teachers fostering IL with their students, our primary goal was to develop a learning design to capably teach IL on the upper secondary level in practice. Furthermore, we wanted to contribute to the theoretical discussion about how to foster IL in upper secondary education. Our main research question derives from these two goals.

> "How to develop and improve a massive open online course (MOOC) to foster information literacy skills in high school (upper secondary education) by means of conjecture mapping?"

The content of the paper is structured as follows: Section 2 elaborates on the rationale of our research project:

1 why a massive open online course (MOOC) is a fitting tool to foster IL

2 why the conjecture map is a favourable tool to design the MOOC; Section 3 explains the applied educational design research (EDR) methodology using a conjecture map; Section 4 provides an overview of the findings of the first research cycle and the adjustments made for the second cycle; Section 5 concludes the paper by discussing the results and providing recommendations for future research.

2 Rationale of the research project

We outlined the importance to foster IL with upper secondary students in the introduction. In the following, we explain, why we do so by developing a MOOC using conjecture mapping.

2.1 Why we developed a MOOC

The rapid technological advances force us to think about new ways how to develop IL and how to support teachers in classroom (Fraillon et al., 2019; Phipps and Lanclos, 2017). Using digital devices and performing online search is an important facet of IL (Horton, 2013). This is the reason for our focus on a digital learning design. Because teachers show insecurities when coping with digital learning designs (Fraillon et al., 2019; Wilson et al., 2019), we provide them with an easy-to-use course. Students should be able to complete the MOOC without technical support. Hence, teachers can focus on supporting the students learning process without worrying about technical challenges.

For IL is considered a set of skills, attitudes and knowledge necessary for students to acquire information in all sorts of topics, the collaboration with teachers is crucial to foster IL (Fraillon et al., 2019). Open educational resources constitute a new digital approach to learning designs that fulfils this need. The openness is crucial to use the potential of digital learning by allowing the interconnected development and the dissemination of new and promising approaches (Dillenbourg, 2016). Open educational resources facilitate new forms of cooperative learning and co-creation-processes. Co-creation-processes involving researchers, teachers, and students, enable the concurrent development, evaluation and improvement of a learning design. The digital learning design allows for the immediate collection of data during the learning process. In this light, design research may be conducted as a collaborative co-creation-process where researchers, teachers and users (e.g., students) improve the learning design by using the collected data (Dillenbourg, 2016). A MOOC provides an open digital learning design facilitating the aforementioned co-creation-process and the therefore required data collection (Deng et al., 2018). Furthermore, a MOOC offers help for teachers to resolve the problem of 'fit' between the required open learning settings and fixed classrooms and timetables.

Major disadvantages of MOOCs should not apply to our learning design. The acceptance of MOOCs, highly depends on the reputation of the institution that created it (Egloffstein et al., 2019). In this context, we may benefit from the standing of the University of St. Gallen (HSG, 2019). A main disadvantage of MOOCs – high dropout rates – should not be a major issue in our case. Firstly, MOOCs with a clear target group – Swiss high school students in our case – show higher completion rates (Paton et al., 2018). Additionally, the interaction with the instructor is decisive for retention (Hone and El Said, 2016). The students are enrolled by their teacher(s) and therefore should in most cases be obliged to finish the course. A MOOC that is combined with in-class interaction (blended learning) provides teachers with a learning design that allows fulfilling the new requirements of an open, learner-centred learning design notwithstanding the fixed timetable (Yousef et al., 2015). Hence, we regard a MOOC to be the ideal learning environment to foster IL.

2.2 Why we used conjecture mapping to develop the MOOC

The EDR method developed by McKenny and Reeves that we applied, consists of three phases that are conducted iteratively (McKenney and Reeves, 2012; Wozniak, 2015):

1. *Analysis and exploration of the learning goals:* In order to be able to plan and design suitable measures for teachers, the required skills initially had to be discussed and formulated (Sandoval, 2014). In a two-year research project with a Swiss school on the upper secondary level we refined the recent IL concepts and measures and formulated the required skills (see Subsection 4.1).
2. *Design and construction of the learning setting:* To develop the MOOC we used conjecture mapping because it is a proven way to design digital learning settings (Wozniak, 2015). Its focus lies on the conjectures (main hypothesis), explicating the set of relationships and thereby facilitating the traceability of the applied rationale (Sandoval, 2014; Wozniak, 2015).
3. *Evaluation and reflection:* The conjectures that lead to the learning design have to be evaluated and, if necessary, to be adjusted (Wozniak, 2015) in order to further improve the MOOC. Moreover, a general insight, concerning the IL learning on the upper secondary level, may be derived and allow to contribute to filling the lack of research in this area.

The conjecture mapping is of great value to the second and third step of the EDR, and has been successfully applied to a similar design research (e.g., Wozniak, 2015). In order to launch an ongoing co-creation-process with teachers and students to further improve the MOOC, the advantage of the conjecture map to explicate the rationale behind the learning setting (Wozniak, 2015), was the main reason for our decision to apply it.

3 Research design and methods

In this section, we outline how we used conjecture mapping in our EDR approach to design a MOOC to foster IL, gain new theoretical insight, and improve the MOOC.

3.1 Educational design research

Design research pursues two main goals; the development of innovative practical solutions and the concurring improvement of theoretical knowledge (Dunning, 2011; Euler, 2011). The methodology of the conducted EDR, followed the three phases of the

EDR-cycle according to McKenney and Reeves (2012), for it has successfully been applied to create digital learning settings using a conjecture map (Wozniak, 2015). The three phases of the EDR-cycle enable a maturing intervention and an improved theoretical understanding (McKenney and Reeves, 2012).

3.2 Analysis and exploration of the learning goals (phase 1)

To define the learning goals, we first had to adapt the concept of IL to schools at the upper secondary level and develop a practicable measure for this concept of IL. In a two-year research project, that we conducted in cooperation with a secondary school in the German speaking part of Switzerland, we developed and tested the '7i framework of IL':

1 To get an overview of existing IL research, a systematic literature analysis has been performed.

2 Based on the literature analysis, and by combining and extending existing models, the 7i framework for measuring IL has been created.

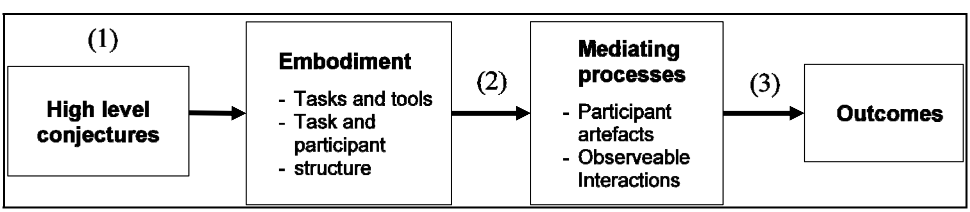
3 The 7i framework for measuring IL has been tested in five classes at a Swiss school on the upper secondary level.

The results of the literature analysis and model development were presented in detail in previous publications. The most important results are outlined in this paper to clarify the learning goals that build the foundation of the designed MOOC (see Subsection 4.1).

3.3 Design and construction of the learning setting (phase 2)

The research of teaching and learning as isolated processes cannot grasp the complexity of the subject (Deng et al., 2018). The conjecture map allows for conducting research that includes the learning and the teaching dimension and its interdependence in a given learning design (Wozniak, 2015) and is therefore a valuable tool to do research in the described way.

Figure 1 The elements of a conjecture map



Source: Adapted from Sandoval (2014)

Conjecture mapping requires the researcher to formulate three different types of conjectures on which the created learning design is based (Figure 1). The *high level conjectures* (1) are the basic assumptions – the applied concept of the learning goals and additionally the understanding of the relevant learning and teaching processes (Sandoval, 2014). Derived from the high level conjectures the *embodiment* of the learning design – materials, tasks, participant structure – is developed. *Design conjectures* (2) reveal how the embodiment should lead to *mediating processes* – observable student artefacts and interactions – that are supposed to induce the learning goals. Finally, the *theoretical conjectures* (3) explain how the mediating processes are supposed to lead to the desired *outcomes* (Sandoval, 2014).

The focus of the conjecture map lies on the different types of conjectures that require studying (Sandoval, 2014) and hence guides phase 3 of the EDR.

The conjecture map was a valuable tool in the applied co-creation-process, for it disclosed the reasoning leading to the learning design. In the development process, the involved teachers, researchers and students could rely on the conjecture map to argue for certain design aspects on its basis or refine the conjecture map if important conjectures were still missing. Every design aspect (e.g., storyline, task types, and levels) that was introduced had to correspond with the current version of the conjecture map. If the aspect did not correspond, the aspect was discarded or the conjecture map had to be adjusted. This way, the conjecture map remained up-to-date and the design was based on theoretical knowledge, since only assumptions that could be proven theoretically were accepted as conjectures. Communication was thereby facilitated and a scientifically-based learning design ensured.

3.4 Evaluation and reflection (phase 3)

In phase 3, the conjectures that led to the learning design were examined (Wozniak, 2015). Do they hold in the constructed learning design? Are there crucial factors missing to explain the functioning of the learning process? How can the learning design be improved to use the theoretical knowledge, represented in the conjecture map?

During the pilot phase from January until June 2019, 632 learners taught in 41 different classes, actively used the developed MOOC. The learners were between 16 and 18 years old and attended an upper secondary school in Switzerland. 23 teachers from 19 different schools enrolled and supervised the participating students. Taking the self-selection-bias (Heckman, 1990) of the involved teachers into account, we checked for previous experience. None of the teachers had already used an online course to teach their students before. Their expertise in handling this type of digital learning settings might therefore represent a common level of Swiss teachers.

In order to evaluate the MOOC and our theoretical understanding represented in the conjecture map, we conducted semi-standardised interviews with the teachers that used the MOOC with their students. In order to counteract a conformity bias, the interviewer used a soft interview-style [Bortz and Döring, (2006), pp.298ff]. Furthermore, he explained that the answers would guide the co-creation-process for the improvement of the MOOC. Thereby the teachers saw the benefit of sharing their critical thoughts and opinions to influence the improvement of the MOOC in a way they desired. Because all of the Interviewees were planning or at least considering to use the MOOC again, this incentive helped to motivate the teachers to disclose all relevant information. To strengthen the reliability, the paraphrased interview was sent back to the interviewee to ensure that his or her answers were completely and correctly recorded and understood. The controlled and adjusted interviews were then analysed to support or challenge the conjectures and to identify propositions to improve the MOOC.

The interview questions were structured along three main topics:

3.4.1 Learning effects

The fostering of IL was the primary goal of our project. We therefore had to find out, if our learning design had a practical impact and increased the IL of the students using it. Thereby we checked for support or doubt about our theoretical conjectures.

3.4.2 Way of using the MOOC

We knew from our previous studies, that many teachers lack the competences for using digital learning settings competently. In order to find out, how the MOOC is used best, we asked the teachers, how they included the MOOC in the curriculum, how they organised and supported their student's learning process with the MOOC and how they would further improve the use of the MOOC. We wanted to know about the challenges and success factors in the work with our MOOC depending on the respective instruction design.

3.4.3 Evaluation of the i-MOOC

To relate the teachers' observations and experiences to our high level conjectures of a good MOOC-design, we asked them to rate the i-MOOC with respect to the criteria defined by Egloffstein et al. (2019). Additionally, we asked for specific improvements they would like us to make to increase the learning effect for the students. Thereby we checked if the embodiment and the design conjectures were plausible.

To capture the *learning effects* and the *way of using the MOOC* we used a set of closed questions, e.g., 'how much did the students IL improve through the use of the i-MOOC?' applying a seven-point Likert scale with extremes 'it improved a lot' to 'it did not improve at all'. To enhance this basic information, we added open follow up questions, e.g., 'in which domains did the students' IL increase?' and 'how did you measure the increase in IL?', to get a deeper understanding of the teachers' experience and actions.

In this contribution we focused on the qualitative evaluation, as the quantitative data that the participants produced while using the MOOC in the pilot phase was not yet available. As soon as the quantitative data is prepared, we can enhance the qualitative evaluation described in this paper and further improve the MOOC and our theoretical understanding.

3.5 Improved theoretical understanding

The practical observations and suggestions of the teachers allowed us to evaluate our conjectures. Which conjectures were supported? Which conjectures were challenged? How should we improve our conjecture map?

3.6 Maturing intervention

The interviews and the improved theoretical understanding helped to identify a range of possible improvements to increase the quality of our MOOC. In line with Plomp (2013) we recognised three main quality dimensions of a learning design: validity, practicality and effectiveness. The improvement process therefore had to strengthen these three quality dimensions of our MOOC.

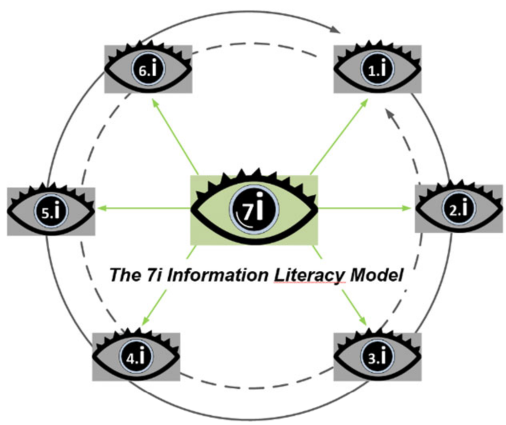
4 Results

First, we present the findings of the three phases of the EDR cycle and then describe how the insights led to a better theoretical understanding and the improvement of the MOOC.

4.1 Phase 1: teaching IL skills in secondary education based on the 7i framework

The starting point for teachers' professional development in the field of IL was to clarify the required learning outcomes. Therefore, we developed an IL framework to provide a basis for teachers on questions such as how to scaffold students' learning processes, give feedback or how to evaluate and grade the IL performance.

Figure 2 7i framework – students' IL (see online version for colours)



The Student who is information literate knows:
(1.i) Information needs: How to determine information needs in the context of a given problem
(2.i) Information sources: Identification of relevant information sources with regard to the identified information needs
(3.i) Information access & seeking strategy: Identification of search and access strategies to access the selected information sources
(4.i) Information evaluation: Evaluation of information sources and resulting information
(5.i) Information use: Appropriate and problem-oriented use of the found information. This might also include understanding in general ethical, legal and socio-political issues of the information or subject- or problem-specific understanding of information
(6.i) Information presentation: Present the information geared to defined target groups
(7.i) Information process & finding reflection: Reflect upon the information search and processing procedure and upon the information resulting from it to learn for future information search processes

The developed 7i framework (Figure 2) contains seven sub-competencies comprising knowledge, skills and attitudes.

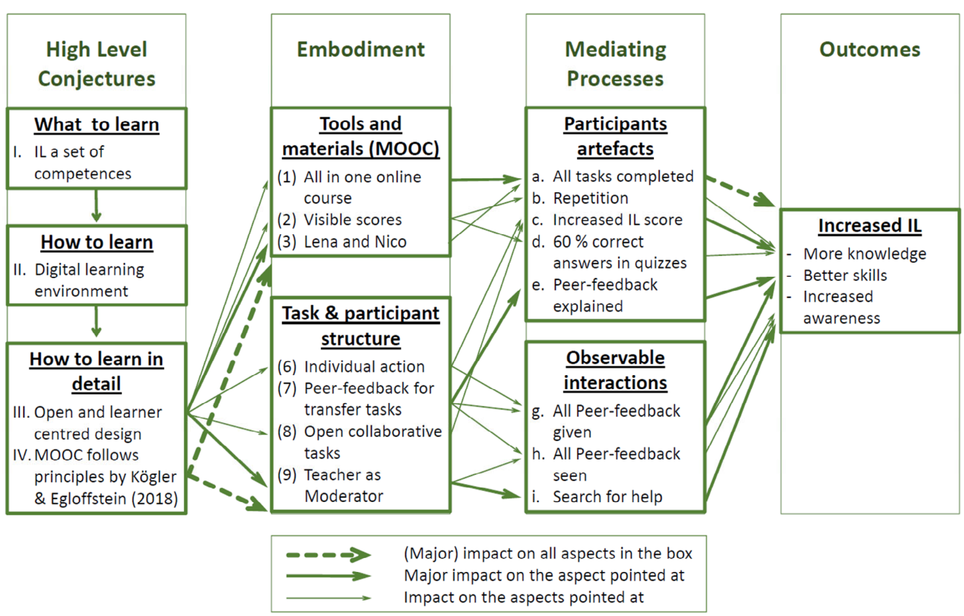
The 7i measurement model results in two different scores of IL: one objective score based on a performance test, and one, rather subjective score, based on self-assessment. The results of the five classes showed a discrepancy between the self-assessed IL score and the objective IL. In the self-assessed test, the students averagely held high scores for all of the seven sub-competences, whereas the objective test unveiled substantial differences among the seven sub-competences. Consequently, a main challenge was to sensitise and support students for the handling of online-based information.

4.2 Phase 2: the initial conjecture map of the i-MOOC

In order to guarantee a protected digital learning space, we use a platform hosted by several Swiss universities (swissmooc.ch). We created the content using the OpenEdx (https://open.edx.org/about-open-edx) software compatible with the platform. We developed the i-MOOC (i-mooc.ch) in a co-creation-process including teachers, scientists and students in order to design a learning setting that covers the need of all the relevant actors and increases the mutual exchange. The result of this co-creation-process is depicted in the following conjecture map.

The conjecture map of the i-MOOC (Figure 3) is explained along four major categories: *high level conjectures*, *embodiment*, *measures*, and *desired outcome*. Additionally, *design conjectures* explain how we expect the embodiment to induce the mediating processes. *Theoretical conjectures* explain why the mediating processes are assumed to induce, the desired outcome.

Figure 3 Conjecture map of the i-MOOC (see online version for colours)



4.2.1 Desired outcome

The desired *outcome* equals in our case the learning goals and therefore entails the increase in IL that the participants should achieve. The knowledge, skills and attitudes defined in the 7i model are the three dimensions of IL competences we want to foster. A detailed rubric of the learning goals is available: https://i-mooc.ch/wp-content/uploads/2020/07/Rubric_LZ_E.xlsx.

4.2.2 High level conjectures

The *high level conjectures* follow from our research and define our understanding of what to learn and how to learn.

a *What to learn:* The content of the i-MOOC bases on the premise that IL consists of a set of competences (I) that include knowledge, skills and attitudes (Euler and Hahn, 2003) which all should be fostered. The interdependence of the competences detected in the 7i model, has a major impact on the i-MOOC design, e.g., the scaffolding of the tasks in order to increase the complexity of the tasks gradually.

b *How to learn:* The second high level conjecture relevant for the i-MOOC entails, that a digital learning environment is crucial to foster IL (II) in theory and practice (Horton, 2013). Considering the poor competences of many teachers to include digital learning designs in their instruction (Fraillon et al., 2019), the i-MOOC is designed for students to master without technical support provided by the teacher.

c *How to learn in detail:* An open and learner centred design is necessary to use the potential of digital learning environments (III) (Dillenbourg, 2016). Even if a MOOC entails the potential to realise these advantages, MOOCs rarely manage to unleash their full potential (Li and Baker, 2018). Deficits may be the lack of authentic problems, activation of participants, and differentiation according to the personal skill levels (Egloffstein et al., 2019; Sinclair and Kalvala, 2016). Therefore, we design our MOOC in correspondence with the principles developed by Egloffstein et al. (2019) (IV):

- *Structure and clarity* – Every module of our MOOC entails an overview containing the structure of the module, the learning goals and the estimated time necessary to complete the module.
- *Merrill criteria* – Together with the protagonists of the i-MOOC, Lena and Nico, the students discover one problem after the other (problem centeredness). The story telling approach, following Nico's and Lena's journey to find out 'why some videos go viral', allows us to demonstrate (demonstration) solutions and encourage reflection of the acquired competences (integration).
- *Further pedagogical principles* – Practical problems solved by Nico and Lena ensure the principle of authenticity. The two are real high school students and thereby help the learner to identify themselves with the protagonists. Besides the created short movies, Nico and Lena also take part as animated characters. Furthermore, students give and receive peer-feedbacks and participate in collaborative exercises.

4.2.3 Embodiment

The conjecture map includes only the tools, tasks, and materials of the i-MOOC, that we directly expect to induce the mediating processes. The vast majority of tasks and materials of the i-MOOC is not part of the conjecture map. These were derived from the high level conjectures too. However, it would exceed this papers' scope to explain every one of them and render the conjecture map too complex to give a good overview (Deng et al., 2018; Sandoval, 2014). You find many examples of the content on https://i-mooc.ch.

4.2.3.1 Tools and materials (i-MOOC)

The all in one online course (1) allows for a clear structure according to the Merrill criteria (Egloffstein et al., 2019) and allows the learner to be flexible in time and place, respecting the needs of a learner centred approach, consistent with high level conjecture III. The personal test scores are visible for the learner (2), not only in the pre-test and posttest, whereby learners are activated (Merrill criterion) and can transparently track their progress (Figure 4).

Figure 4 Scoreboard i-MOOC (see online version for colours)

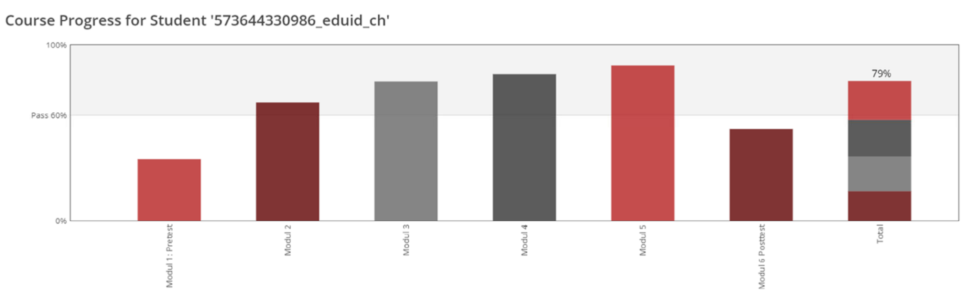


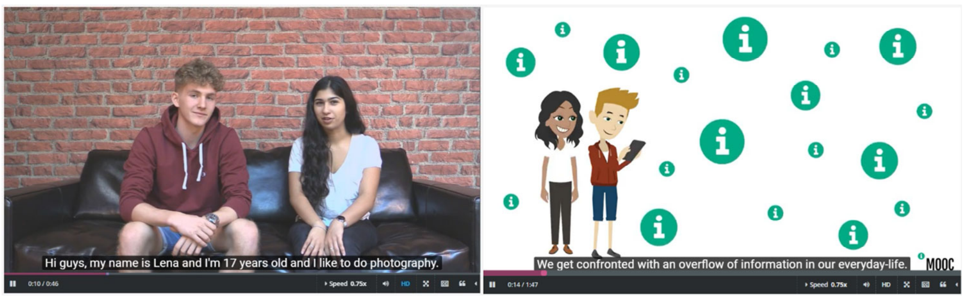
Figure 5 Course structure and levels (see online version for colours)

Course Structure and Levels		
Module	**Content**	**Level**
M1	• **Pretest** • **Introduction**	**Info Tourist**
M2	• Define the information need • Chose the information source	Info Mapper
M3	• Search relevant information • Evaluate information	Info User
M4	• Use information competently • Present information adequately	Info Guide
M5	• Internet Rally	Info Leader
M6	• Posttest • Wrap up	Info Master

The scores in combination with the levels and badges, may as game elements also rise the motivation (Jones-Kavalier and Flannigan, 2016) to advance step by step from ‘info tourist’ to ‘info master’ (Figure 5).

Among the most important decisions was the choice of the story telling approach with two actual high school students, Lena and Nico (3), as the main characters guiding the students through the course. They appear in 8 short movies, that tell the main story, as actors and in an animated form in the 22 instruction videos (Figure 6).

Figure 6 Lena and Nico (see online version for colours)



They underline the learner centrism and support the peer-learning, additionally implemented in the collaborative tasks and peer-feedbacks.

4.2.3.2 Task and participant structure

The combination of individual tasks (4), peer-feedbacks (5) and open collaborative tasks allows the asynchronous tackling of tasks as much as social learning, and thereby creates a flexible course for the participants. The peer-feedback is applied in the transfer tasks. In order to induce integration (Merrill criterion), the learners reflect on their newly acquired skills by discussing their challenges and solutions with classmates and teacher(s). Moreover, they perform a substantial number of peer-feedback tasks where they are able to perform social learning. The teacher’s role as moderator (6) is decisive in this flexible and asynchronous learning setting. As the students have the possibility to adapt according to their own capacity, the teacher should be the one to assure the professional embedding of the process, to guarantee sustainable learning.

4.2.4 Mediating processes

The students may complete a major part of the i-MOOC aside from the regular schedule. This flexible and learner centred design has one major disadvantage regarding the EDR. Many actions are not observable for the research team. However, the research team interviews the teachers participating in the pilot phase of the i-MOOC, to reveal interactions between the teachers and the students. The collection and analysis of the students’ artefacts is facilitated by the digital learning environment and allows to capture the mediating processes.

- *Participants artefacts:* To benefit from the learning design, the students should complete all tasks (a). Considering that MOOCs suffer from high dropout rates (Liyanagunawardena et al., 2014), besides the all in one online course (1), the school setting should prevent a high number of dropouts. However, the responsibility for taking the whole course lies, respecting the learner-centred approach, with the students. Lena and Nico as well as the visibility of the test scores should help to incentivise the students to complete all tasks and thereby prevent the lack of motivation often seen concerning IL (Beutelspacher, 2014). Moreover, the content is straightforward and the storytelling may induce curiosity similar to a telenovela, where the viewers want to know how it continues. The students take an IL test at the beginning and at the end of the i-MOOC. Hence, might become aware of their learning potential and their progress is honoured. The diverse tasks including a lot of practical application and reflection in the transfer tasks, should lead to a higher IL score (c) and therefore to an increased IL. The validated IL test (Balceris, 2011) has in our previous studies proven to measure the increase in IL. The visibility of the scores and the possibility to repeat quizzes and tasks, should lead to at least 60% correct answers in the quizzes (d), as we defined this threshold to pass the course and students are likely to be motivated by this game-element (Jones-Kavalier and Flannigan, 2016) of collecting points to succeed. Repeating the quizzes strengthens and evaluates the IL knowledge alike. If students add reasonable explanations to their peer-feedback, they not only allow the peer to learn better but demonstrate higher involvement that may lead to a better skills and an increased awareness (Hattie and Timperley, 2007). The artefacts are collected in the i-MOOC and therefore facilitated an efficient handling and evaluating in the latter phase of the EDR.
- *Observable interactions:* The exchange with peers in the open collaborative tasks and the peer-feedbacks (f, g) build the main part of the interactions during the MOOC. The more the students appreciate and therefore conduct the peer-feedback the more likely it is for them to increase their IL skills and awareness (Hattie and Timperley, 2007). The seeking for help (h) indicates the awareness of personal IL deficits, for poor performers often do not realise their own deficits in a specific field (Dunning, 2011). Therefore, it might be evidence for an increased awareness for the challenges of IL.

4.3 Phase 3: evaluation and reflection of the pilot phase

Sixteen teachers that used the i-MOOC with their students during the six-month pilot phase were interviewed. They supervised between one and five classes including 515 out of 632 students who actively used the i-MOOC. Two of the 16 teachers we interviewed, stopped using the i-MOOC as they were not satisfied with the learning design. Their critical feedback seems to be particularly interesting to improve the learning design.

4.3.1 Learning effects

All the teachers whose students were actively using the i-MOOC attributed their students an increase in IL. Six attributed a strong learning effect to the i-MOOC, eight saw small improvements. The main learning effects were identified to be the improved use of searching tools (skill), the definition of the problem (skill) and the general understanding of IL and its importance (knowledge and awareness). Four teachers emphasised that especially the high activity level of the students, required in the i-MOOC, led to an increase in IL. Hence, the designed MOOC is an effective tool to teach IL.

For the further improvement of the i-MOOC we need to know which competences improved and which competences did not improve. According to multiple teachers, their students got better in defining a research question (seven teachers) and conducting a refined online research (ten teachers) to find relevant information (five teachers). Besides, four classes' general awareness for the importance of IL increased. Nevertheless, no teacher observed better presentation skills (6th i of the 7i-model) with their students. Maybe their observations that based on written tests, written school assignments or discussions with the students missed an existing increase in presentation competences. More likely, the i-MOOC does not deliver enough support and guidance in that area and needs improvement.

4.3.2 Way of using the i-MOOC

- *Inclusion in curriculum:* The teachers connected the i-MOOC to various topics when introducing the i-MOOC to their students. Whereas some stressed the importance of IL for writing school assignments, others mainly pointed out the enormous information exposure and the recent challenges regarding fake news, to illustrate the relevance of the i-MOOC to their students. Notwithstanding the different approaches and expectations from the teachers, there surfaced no differences in learning effects caused by the differing approaches and expectations could be identified.
- *Organisation of the learning process:* The teachers organised the learning with the i-MOOC diversely. One major difference was the level of support and guidance that the teachers provided. Six followed their students' activity in the i-MOOC closely and supported their learning process through intensive support, supervision, discussions in class, or supplementary tasks. Nine did not monitor the students' actions in the i-MOOC, only discussed few elements in class, and only provided support when actively approached by the students. In addition to the supervision of the learning process in the i-MOOC, some teachers incentivised their students to engage intensively with graded tests about IL and the i-MOOC (five teachers) or graded some of the tasks the students solved in the i-MOOC (two teachers). Closer supervision or grading has led to a higher learning effect. The decisive aspect here seems to be the obligation created for the learners.
- *Reflection on improving the use:* Most teachers plan to take a more active role when using the i-MOOC again. More discussions to deepen the course content in class (a stronger) monitoring of the learning process of their students, and a graded test were named by several teachers as planned improvements. Only two of the interviewed teachers were entirely satisfied with their use and plan to use the i-MOOC the same way again – both supported their students intensively and used graded tests as further incentives.

4.3.3 Evaluation of the i-MOOC

- *Identified strengths:* Most teachers liked the all in one online course approach, because the course was easy to use with their classes. They saw few technical difficulties and stressed the value of the intensive and quick technical support that helped them overcome these difficulties. Furthermore, the clear structure and precisely defined learning goals made the work with the i-MOOC uncomplicated. The underlying 7i-model was understandable and easily applicable. Some teachers created tests and further tasks using this concept of IL. Besides, most teachers identified the high level of activation and transfer as a key success factor for the improved IL. The students could not just 'consume' the course but had to read longer texts, take many quizzes, conduct several research queries, produce written answers and visualisations, and deliver feedback for their peers' solutions. The collaborative tasks and the peer-feedbacks were appreciated by several teachers for having a motivating effect. The teachers also liked the story-telling approach with Nico and Lena (see Subsection 4.2) even if two of them observed that their students could not relate to the characters and the story. The gamification elements (points, levels, etc.) showed a mixed result. While some classes responded very well to the possibility to collect points and therefore repeated quizzes to improve their score, others were frustrated because the scoring system was not always transparent. Orji (2014) suggests that a tailored gamification experience might improve the motivational effect for certain user types.

- *Identified weaknesses:* The two teachers who stopped working with the i-MOOC had one major complaint. Both considered the login process to be too time-consuming and the course to be less efficient than traditional in-class-instruction. One of them also mentioned, that the difficulty level of the i-MOOC was not demanding for his class. All the other teachers rated the i-MOOC as a valuable tool to instruct IL. However, several also asked us to improve the efficiency of the course design. Especially the login process should be simplified. Many teachers also asked us to shorten the learning videos and lessen redundancies and repetition. Three teachers, who criticised the efficiency, proposed a modular course to simplify a selective use of the content. Furthermore, we need to improve the peer-feedback tasks. Many teachers reported, that their students complained about the poor quality of feedbacks they received and the poor quality of answers they had to feedback. However, most of the teachers consider the peer-feedback tasks to be a valuable part of the i-MOOC.

  The i-MOOC does not guide the students to take notes. Many teachers did not support their students and were surprised that they did not take appropriate notes. They would expect the i-MOOC to provide stronger help to take notes and offer a possibility to download summaries. Also, to conduct a graded test or to extend the course additional material was demanded by four of the interviewed teachers. Only one of these teachers designed an own test and some extra-tasks to enhance the i-MOOC, the others expected the i-MOOC to provide such material for them.

  Even if most teachers liked the gamification concept basing on the progression from info-tourist to info-master and the attribution of points for the quizzes, transfer tasks, and peer-feedbacks, it has some flaws. The teachers criticised that some of the students only went to collect the points instead of focusing on the content. Others were mainly critical about the non-transparent scoring in the pre-test and post-test.

- *Validity, practicability, effectiveness:* The validity of the i-MOOC is supported by the teachers. They see, however, potential to increase the practicability and effectiveness of the learning design through an improvement of the course design and a better instruction by themselves.

4.4 Improved theoretical understanding

Figure 7 summarises the theoretical findings concerning the existing conjectures. Red highlighted (darker) conjectures were questioned; green highlighted (lighter) conjectures were supported by the teachers. The other conjectures (not highlighted) could not be rated conclusively and need to be revisited when the quantitative data is available.

Figure 7 Design conjectures and theoretical conjectures (see online version for colours)

Design Conjectures

DC1	all in one course	–	all tasks completed
DC2	visible scores	–	repetition
DC3	visible scores	–	60% correct
DC4	Lena and Nico	–	all tasks completed
DC5	individual action	–	increased IL score
DC6	individual action	–	all peer-feedback given
DC7	peer-feedback for transfer tasks	–	peer-feedback explained
DC8	peer-feedback for transfer tasks	–	all peer-feedback given
DC9	peer-feedback for transfer tasks	–	all peer-feedback seen
DC10	open collaborative tasks	–	increased IL score
DC11	teacher as moderator	–	all peer-feedback seen
DC12	teacher as moderator	–	search for help

Theoretical Conjectures

TC1	all tasks completed	–	more knowledge
TC2	all tasks completed	–	better skills
TC3	all tasks completed	–	increased awareness
TC4	repetition	–	more knowledge
TC5	increased IL score	–	more knowledge
TC6	60% correct	–	more knowledge
TC7	peer-feedback explained	–	better skills
TC8	all peer-feedback given	–	better skills
TC9	all peer-feedback given	–	increased awareness
TC10	all peer-feedback seen	–	increased awareness
TC11	search for help	–	increased awareness

4.4.1 Supported conjectures

The *all in one course* (DC1) including *visible scores* (DC2) as gamification elements and a *story telling approach* (DC4) seem to contribute to a stronger engagement of the students, hence, more *repetition* (TC4), what increases the knowledge of the students according to their teachers. Furthermore, the *role of the teacher* (DC11–12; TC11) is decisive for the professional embedding of the learning process and the completion of the tasks in the i-MOOC and subsequently for the learning success.

4.4.2 Challenged conjectures

Even if the inclusion of *peer-feedback* (DC6–9; TC7–10) tasks in the i-MOOC was appreciated by the teachers, in the current form these tasks do not guarantee a contribution to the learning success. Only if the peer-feedback is conducted seriously, it contributes to a higher IL. Regular MOOCs often show challenges with peer-feedback quality (Luo et al., 2014). We hoped, that in a school setting these problems would not appear, but apparently this was only the case for some classes. For many teachers the underlying transfer tasks were much more important than the feedback-process.

The teachers suggested that the all in one course was helpful for the practicability but did not necessarily lead the students to *complete all the tasks* (TC1–3). Several teachers left it to the students whether they wanted to complete all the tasks of the i-MOOC. They do not agree with our assumption that the completion of all tasks leads to an increased IL. An efficient learning process that includes skipping elements that seem already familiar, was favoured by the teachers.

4.4.3 New conjectures to be considered

According to the teachers, an *efficient learning design* is of the utmost importance for their decision to use the course and to keep the students active and motivated. An *active role of the teacher* seems to be decisive for the learning success. If the teacher does not motivate, incentivise, or support the students enough, the i-MOOC suffers from the same deficiencies as many other MOOCs. Hence, the i-MOOC needs to support the teachers more in the organisation of the learning setting and the creation of obligation for the learners, as a complete course seems not to be enough.

4.4.4 New conjecture map

Figure 8 depicts the new conjecture map, including the changes described in the previous section (changed aspects marked with the diamond icon). The main changes concern the peer-feedbacks, the active role of the teacher and the efficiency of the learning design. The peer-feedbacks have to be of good quality to cause rising competences. Teachers actively engaging in the learning process through support and further discussions, helped the students learning success and motivation.

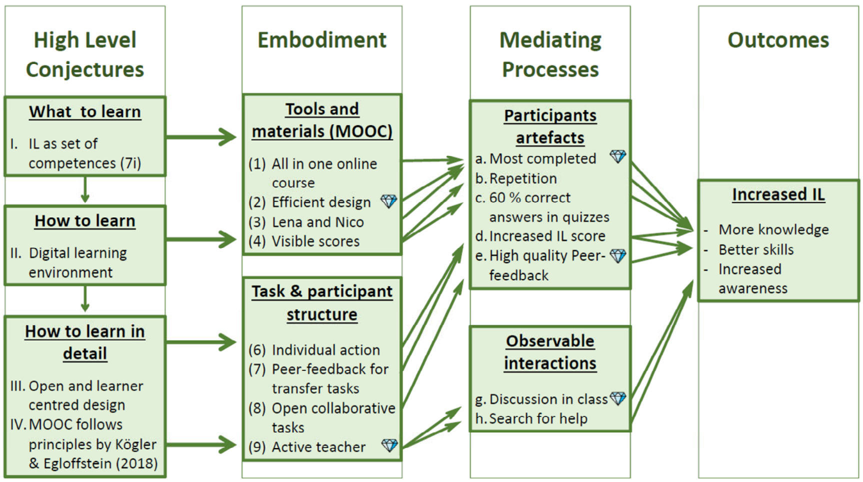
Furthermore, we have no reason to assume that some conjectures are more important than others, so we withdrew this differentiation in the new conjecture map.

4.4.5 Learnings of the work with the conjecture map

The conjecture map helped in the co-creation-process to steadily communicate the state of thought and keep everyone working on the i-MOOC updated on the relevant ideas, leading to the current state of the learning design. The theoretical rooting was thereby facilitated.

In the evaluation phase of the EDR process, the conjecture map guided the assessment design. The interviews, however, showed the importance to conduct a wider evaluation. Because the conjecture map reduces the whole project to its (considered) core factors and assumptions, a large part of the learning design and of the reasoning drops out. Only with a more general evaluation approach we were able to find entirely new aspects arising from the practical use of the i-MOOC to enhance our theoretical understanding and thereby improve the learning design.

Figure 8 New conjecture map of the i-MOOC (see online version for colours)



4.5 *Maturing intervention*

The most important goal of our research project was to develop and improve a learning design to foster IL. In the current improvement process we therefore strive to improve its validity, practicability, and effectiveness (Plomp, 2013). Our improved theoretical and practical understanding of the i-MOOC unveiled several elements that need improvement:

1 *The login process* unfortunately cannot be simplified. In order to guarantee a protected digital learning space, we use a platform hosted by several Swiss universities. As the secondary students, using the i-MOOC, do not have a digital university-student-identity (edu-ID) yet, the login process first includes the generation of this edu-ID. Without more funding it is impossible for us to change the host or host the i-MOOC ourselves and still guarantee the same level of data protection. Therefore, we will offer more support during the registration period and better sensitise the teachers to the task in advance.

2 *The efficiency* of the learning process, has to be improved. Even if the MOOC criteria (Egloffstein et al., 2019) we used as a guideline stress the importance of a clear structure and repeated mentioning of the learning goals, we might reduce redundancies by shortening the video material and leaving only the written guidance. However, we will not cut content repetition, to give the students an opportunity to repeat and enhance certain aspects.

3 *The peer-feedback* tasks need to be better introduced and guided. Studies show, that providing a grading rubric and sample answers to the peer-feedback tasks (Meek et al., 2017; Kulkarni et al., 2013), might improve the feedback quality.

4 Maybe the most important task will be to *support the teachers* even more in working with the i-MOOC. Based on this evaluation, we can provide additional material to enhance the course and illustrate promising ways to organise the learning processes. We certainly will keep up the already in place instructional webinars and the offered support desk.

5 We will support *students' note taking* by providing structured templates, the students can complete with their findings. This should also contribute as a means of guidance for the teachers. Apparently, most of them left the note-taking completely to the students and realised later that most students did not take any notes.

6 In order to *increase the transparency* for the pre-test and posttest, we will deliver the results in pieces and add written explanations. We cannot show the solutions to save the integrity of the test.

The evaluation showed no ground to question the validity of the i-MOOC. With the outlined measures, the practicability and effectiveness of the i-MOOC should further increase and improve the constructed learning design.

5 Discussion and future perspectives on developing IL in an open digital learning setting

This paper addresses the problem of 'fit' between new requirements in terms of digital skills such as IL and prevailing conditions of the schooling system – classroom and fixed timetables. The need for open digital learning settings, that include a broad number of participants and allows for co-creation of the learning setting (Dillenbourg, 2016), is a challenge. Especially in the light of teachers' lack of competences concerning the use and development of digital learning settings (Fraillon et al., 2019), supporting tools have to be developed. The i-MOOC fulfils this requirement as it provides teachers with a flexible and easy-to-use tool to foster IL in an open learning setting that allows for co-creation-processes.

The evaluation indicates, that the i-MOOC is an effective tool to foster IL and that the active management of the learning processes by the teachers had a significant impact on the learning outcomes. Although the involved teachers never used online courses with their students before, the fostering of IL with help of a MOOC appears to work. With more experience, the teachers are likely to improve their abilities (Passarelli and Kolb, 2011) in guiding the students' learning process and facilitate higher learning outcomes. Furthermore, the efficient design of a MOOC appears to be an important factor to motivate teachers on the upper secondary level to use it.

We are aware, that the described EDR focuses on fostering IL in the Swiss school system. Not all aspects of the conjecture map may be transferrable to other contexts. Yet, the conjecture map proved to be a valuable tool for the theory led construction and the focused evaluation of the i-MOOC. Especially the second and third phase of the EDR were facilitated by the conjecture map. Analysing the quantitative data will be the next step of the EDR process (McKenney and Reeves, 2012; Wozniak, 2015) to refine the i-MOOC and enhance the theoretical understanding and improve the learning design.

On the way to an open educational ecosystem (Dillenbourg, 2016), the developments in learning analytics will be decisive to allow for adaptive learning processes using the generated analytical data and hence allow for personalised learning designs (Cope and Kalantzis, 2016). Besides, the necessary competences for teachers and students to handle information and communication technologies (ICTs) will change further. Computational thinking will be key to enhance the interaction and competent application of computers (Fraillon et al., 2019) and therefore a prerequisite for IL.

References

Argelagós, E. and Pifarré, M. (2012) 'Improving information problem solving skills in secondary education through embedded instruction', *Computers in Human Behavior*, Vol. 28, No. 2, pp.264–270, DOI: 10.1016/j.chb.2011.10.024.

Balceris, M. (2011) *Medien- und Informationskompetenz: Modellierung und Messung von Informationskompetenz bei Schülern*, PhD thesis, Freie Universität Berlin, Berlin, Germany.

Beutelspacher, L. (2014) 'Assessing information literacy: creating generic indicators and target group-specific questionnaires', *European Conference on Information Literacy*, Springer, Cham, Vol. 492, pp.521–530.

Bortz, J. and Döring, N. (2006) *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*, 4th ed., Springer-Medizin-Verlag, Heidelberg.

Coklar, A.N., Yaman, N.D. and Yurdakul, I.K. (2017) 'Information literacy and digital nativity as determinants of online information search strategies', *Computers in Human Behavior*, Vol. 70, pp.1–9, DOI: 10.1016/j.chb.2016.12.050.

Cope, B. and Kalantzis, M. (2016) 'Big data comes to school: implications for learning, assessment, and research', *AERA Open*, Vol. 2, No. 2, pp.1–19, DOI: 10.1177/2332858416641907.

Deng, R., Benckendorff, P. and Gannaway, D. (2018) 'Progress and new directions for teaching and learning in MOOCs', *Computers & Education*, Vol. 129, pp.48–60, DOI: 10.1016/j.compedu.2018.10.019.

Dillenbourg, P. (2016) 'The evolution of research on digital education', *International Journal of Artificial Intelligence in Education*, Vol. 26, No. 2, pp.544–560, DOI: 10.1007/s40593-016-0106-z.

Dunning, D. (2011) 'The Dunning-Kruger effect: on being ignorant of one's own ignorance', *Advances in Experimental Social Psychology*, Vol. 44, pp.247–296, DOI: 10.1016/B978-0-12-385522-0.00005-6.

Egloffstein, M., Koegler, K. and Ifenthaler, D. (2019) 'Instructional quality of business MOOCs: indicators and initial findings', *Online Learning: OLJ = Online Learning Journal*, Vol. 23, No. 4, pp.85–105.

Euler, D. (2011) 'Wirkungs- vs. gestaltungsforschung – eine feindliche Koexistenz?', *Zeitschrift für Berufs- und Wirtschaftspädagogik*, Vol. 107, No. 4, pp.520–542.

Euler, D. and Hahn, J. (2003) *Wirtschaftsdidaktik*, Haupt, Berlin.

Fraillon, J. et al. (2019) *Preparing for Life in a Digital World: The IEA International Computer and Information Literacy Study 2018 International Report*, Amsterdam, IEA.

Gapski, H. and Tekster, T. (2009) *Informationskompetenz in Deutschland – Überblick zum Stand der Fachdiskussion und Zusammenstellung von Literaturangaben, Projekten und Materialien zu einzelnen Zielgruppen*, Landesanstalt für Medien Nordrhein-Westfalen, Düsseldorf.

Hartig, J. and Jude, N. (2007) 'Empirische erfassung von kompetenzen und psychometrische kompetenzmodelle', in Hartig, S. and Klieme, E. (Eds.): *Möglichkeiten und Voraussetzungen Technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung*, pp.17–36, Bundesministerium für Bildung und Forschung, Berlin.

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of Educational Research*, Vol. 77, No. 1, pp.81–112, DOI: 10.3102/003465430298487.

Heckman, J.J. (1990) 'Selection bias and self-selection', in *Econometrics*, pp.201–224, Palgrave Macmillan, London.

Hone, K.S. and El Said, G.R. (2016) 'Exploring the factors affecting MOOC retention: a survey study', *Computers & Education*, Vol. 98, pp.157–168.

Horton, F.W. (2013) *Overview of Information Literacy Resources Worldwide*, UNESCO, Paris.

HSG (University of St. Gallen) [online] https://www.unisg.ch/en/universitaet/ueber-uns/portraet/rankingsundakkreditierungen/rankingresultate (accessed 3 January 2019).

Jones-Kavalier, B. and Flannigan, S.L. (2006) 'Connecting the digital dots: literacy of the 21st century', *Educause Quarterly*, Vol. 29, No. 2, pp.8–10.

Koltay, T., Spiranec, S. and Karvalics, L.Z. (2016) *Research 2.0 and the Future of Information Literacy*, Chandos Publishing, Cambridge.

Kulkarni, C., Wei, K.P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D. and Klemmer, S.R. (2013) 'Peer and self assessment in massive online classes', *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 20, No. 6, pp.33:01–33:33.

Lanning, S. and Mallek, J. (2017) 'Factors influencing information literacy competency of college students', *The Journal of Academic Librarianship*, Vol. 43, No. 5, pp.443–450.

Li, Q. and Baker, R. (2018) 'The different relationships between engagement and outcomes across participant subgroups in massive open online courses', *Computers & Education*, Vol. 127, pp.41–65.

Liyanagunawardena, T., Parslow, P. and Williams, S. (2014) 'Dropout: MOOC participants' perspective', *EMOOCs 2014, The Second MOOC European Stakeholders Summit*, pp.95–100.

Luo, H., Robinson, A. and Park, J.Y. (2014) 'Peer grading in a MOOC: reliability, validity, and perceived effects', *Online Learning Journal*, Vol. 18, No. 2, pp.1–14.

McKenney, S. and Reeves, T. (2012) *Conducting Educational Design Research*, Routledge, Oxon.

Meek, S.E., Blakemore, L. and Marks, L. (2017) 'Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review', *Assessment & Evaluation in Higher Education*, Vol. 42, No. 6, pp.1000–1013.

Ng, W. (2012) 'Can we teach digital natives digital literacy?', *Computers & Education*, Vol. 59, No. 3, pp.1065–1078.

Orji, R. (2014) *Design for Behaviour Change: A Model-Driven Approach for Tailoring Persuasive Technologies*, PhD thesis, University of Saskatchewan, Canada.

Passarelli, A.M. and Kolb, D.A. (2011) 'The learning way: learning from experience as the path to lifelong learning and development', in London, M. (Ed.): *The Oxford Handbook of Lifelong Learning*, 1st ed., pp.70–90, Oxford University Press, New York.

Paton, R.M., Scanlan, J.D. and Fluck, A.E. (2018) 'A performance profile of learner completion and retention in Australian VET MOOCs', *Journal of Vocational Education & Training*, Vol. 70, No. 4, pp.581–599, DOI: 10.1080/13636820.2018.1463278.

Phipps, L. and Lanclos, D.M. (2017) 'Leading with digital in an age of supercomplexity', *Irish Journal of Technology Enhanced Learning*, Vol. 3, No. 1, pp.1–11, DOI: 10.22554/ijtel.v2i2.22.

Plomp, T. (2013) 'Educational design research', in Plomp, T. and Nieveen, N. (Eds.): *An Introduction. Educational Design Research*, pp.11–50, Netherlands Institute for Curriculum Development, Enschede.

Prensky, M. (2011) *Digital Natives, Digital Immigrants*, Vol. 9, No. 5, pp.1–6, MCB University Press, DOI: 10.1108/10748120110424816.

Sandoval, W. (2014) 'Conjecture mapping: an approach to systematic educational design research', *Journal of the Learning Sciences*, Vol. 23, No. 1, pp.18–36, DOI: 10.1080/10508406.2013.778204.

Sinclair, J. and Kalvala, S. (2016) 'Student engagement in massive open online courses', *International Journal of Learning Technology*, Vol. 11, No. 3, pp.218–237.

Šorgo, A., Bartol, T., Dolničar, D. and Boh Podgornik, B. (2017) 'Attributes of digital natives as predictors of information literacy in higher education', *British Journal of Educational Technology*, Vol. 48, No. 3, pp.749–767.

Wilson, A., Fuhrman, O. and Turner, K. (2019) ''Digital tools will never take the place of a good teacher': understanding teachers' resistance to using technology through Glasser's choice theory', *International Journal of Learning Technology*, Vol. 14, No. 1, pp.42–58.

Wozniak, H. (2015) 'Conjecture mapping to optimize the educational design research process', *Australasian Journal of Educational Technology*, Vol. 31, No. 5, pp.597–612.

Yousef, A.M.F., Chatti, M.A., Schroeder, U. and Wosnitza, M. (2015) 'A usability evaluation of a blended MOOC environment: an experimental case study', *The International Review of Research in Open and Distributed Learning*, Vol. 16, No. 2, pp.69–93, DOI: 10.19173/irrodl.v16i2.2032.

Studie 4:

Guggemos, J., Moser, L. & Seufert, S. (2022). Learners don't know best: Shedding light on the phenomenon of the K-12 MOOC in the context of information literacy. *Computers & Education*, 188, 104552.

Computers & Education 188 (2022) 104552

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Learners don't know best: Shedding light on the phenomenon of the K-12 MOOC in the context of information literacy

Josef Guggemos[a,*], Luca Moser[b], Sabine Seufert[b]

[a] University of Education Schwäbisch Gmünd, Department of Vocational Education and Training, Oberbettringer Str. 200, 73525, Schwäbisch Gmünd, Germany

[b] University of St.Gallen, School of Management, Institute for Educational Management and Technologies, St. Jakob-Strasse 21, 9000, St. Gallen, Switzerland

ARTICLE INFO

Keywords:
K-12 MOOC
Blended MOOC
Upper secondary level education
Technology acceptance
Learning effects

ABSTRACT

Massive Open Online Courses (MOOCs) have received much attention in higher education; however, evidence about MOOCs at the K-12 level is scarce. To shed light on the phenomenon, we use the i-MOOC that aims at fostering upper secondary level students' information literacy. The i-MOOC is a blended MOOC developed and refined in a design research process; it meets established criteria for high-quality MOOCs. In 2020, 1032 upper secondary level students in German-speaking Switzerland took the i-MOOC; the sample comprises $N = 167$ students who voluntarily filled in a questionnaire. The students are mainly from high schools and vocational schools. Learning effects are captured with a performance test. Information literacy gains are significant and medium in size: $d = 0.75$. The technology acceptance of students is evaluated using the extended unified theory of acceptance and use of technology (UTAUT2). Student technology acceptance of K-12 MOOCs is primarily driven by hedonic motivation, i.e., perceived fun and entertainment. However, this type of motivation negatively predicts learning gains. Implications for teachers and educational decision makers are discussed.

1. Introduction

Massive Open Online Courses (MOOCs) are a much debated phenomenon in tertiary education (L. Li, Johnson, Aarhus, & Shah, 2022; Reich & Ruipérez-Valiente, 2019). They are built to be attended by a large number of students (massive) with unrestricted access (open), and to be performed on the web (online) within a self-sufficient learning environment (course). Research activity on MOOCs have experienced a sharp increase (Joksimović et al., 2018; Zhu, Sari, & Lee, 2020), during the course of which several challenges have been identified. One of the most prominent is the notoriously high dropout rate (Hone & El Said, 2016; Paton, Scanlan, & Fluck, 2018).

In light of such challenges, the integration of MOOCs into formal learning settings has gained attention, i.e., MOOCs are integrated into in-class courses (Meet & Kala, 2021; Yousef & Sumner, 2021; Zhu, Sari, & Lee, 2020). Such *blended MOOCs* aim to combine the advantages of in-class and online learning (Chen, Zou, Xie, & Zou, 2020; Israel, 2015; Wang, Hall, & Wang, 2019; Zhu, Sari, & Lee, 2020). Blended MOOCs comprise in-class learning enriched with available traditional MOOCs, as well as MOOCs specifically designed for blended learning purposes. Both settings might be valuable if appropriately implemented (Bralić & Divjak, 2018). In traditional MOOCs, students usually lack sufficient guidance and support (Kasch, Van Rosmalen, & Kalz, 2021), and struggle with self-regulation

* Corresponding author.
E-mail address: josef.guggemos@ph-gmuend.de (J. Guggemos).

https://doi.org/10.1016/j.compedu.2022.104552
Received 16 February 2022; Received in revised form 17 May 2022; Accepted 19 May 2022
Available online 1 June 2022

(Jansen, van Leeuwen, Janssen, Conijn, & Kester, 2020; Kizilcec, Pérez-Sanagustín, & Maldonado, 2017). In blended MOOCs, lecturers could provide elaborate feedback to students, for instance, using the functionalities of learning analytics (Lu et al., 2018). Available evidence points to positive engagement and learning effects of blended MOOCs (K. Wang & Zhu, 2019). Moreover, high dropout rates may not be a problem in blended MOOCs (Cohen & Magen-Nagar, 2016; Muñoz-Merino et al., 2013). Against this backdrop, it is not surprising that blended MOOCs are increasingly being implemented (de Moura et al., 2021); however, there may also be downsides to blended MOOCs. They are more resource intensive to scale for a massive number of students than traditional MOOCs as lecturers are necessary. Furthermore, designing a high-quality blended MOOC is challenging and learning outcomes might be poor if the blended MOOC were inappropriately used by lecturers (Bralić & Divjak, 2018; de Moura et al., 2021).

Despite the widespread attention given to (blended) MOOCs in higher education, there is little evidence about K-12[1] MOOCs. The literature review of Koutsakas, Chorozidis, Karamatsouki, and Karagiannidis (2020), covering the years 2013–2020, identified only twenty-one studies on K-12 MOOCs. Most of these address STEM-related topics. For instance, the Department of Computer Science at the University of Helsinki offers a computer science MOOC that can be taken by any student in any Finnish school (Kurhila & Vihavainen, 2015).

Evidence from higher education may not be directly helpful to gain an understanding of the phenomenon K-12 MOOCs: K-12 learning processes are different from those at the tertiary level (Briggs & Crompton, 2016; Koutsakas, Chorozidis, et al., 2020). Moreover, it may be suboptimal to use MOOC content tailored to tertiary education at the K-12 level: high quality instruction should consider learners' prior knowledge and background (Bransford, Brown, & Cocking, 2000).

The aim of the paper at hand is to shed light on the phenomenon of K-12 MOOCs by providing quantitative evidence from an information literacy blended MOOC offered in German-speaking Switzerland to upper secondary level students. The project was funded by the Swiss National Science Foundation. We make three contributions to the literature. First, based on a review of the literature, we present a conceptual framework showing how K-12 MOOCs can be evaluated. Second, to our knowledge, we are the first study to provide quantitative empirical evidence about K-12 MOOCs that goes beyond univariate analyses. Third, we discuss the criteria that should be taken into account when deciding whether to integrate a MOOC into K-12 education.

2. Review of the literature

2.1. Search strategy

The recent literature review of Koutsakas, Chorozidis, et al. (2020) on K-12 MOOCs covers the period from January 2013 to March 2020. We utilized the search strategy of these authors to identify further studies published between March 2020 and December 2021, including ones in German. From the twenty-one studies reviewed by Koutsakas, Chorozidis, et al. (2020), we excluded two. They do not provide empirical evidence (Dziabenko & Tsourlidaki, 2018) or do not follow our conception of a MOOC (Canessa & Pisani, 2013). The latter investigated the delivery of recorded lectures, which is not a self-sufficient learning environment, i.e., does not meet our definition of a course. On the other hand, we included two studies published in 2020 and 2021, and three studies published in German. Overall, we identified twenty-four studies that:

- address MOOC(s) in a K-12 setting,
- are published in peer-reviewed journals or in peer-reviewed conference proceedings,
- report empirical evidence based on primary research,
- are published between January 2013 and December 2021,
- are in English or German.

The supplementary material S1 shows our identification strategy, the reviewed studies, and summarizes the evaluation approaches and main findings of the studies.

2.2. Characteristics of K-12 MOOCs

As Table 1 illustrates, the MOOCs in the reviewed studies can be classified according to three criteria: 1) blended MOOC (yes/no), 2) designed for K-12 (yes/no), and 3) the topic of the MOOC.

Blended MOOCs are integrated in a formal school setting and combine (= blend) both online and in-class learning activities (Bralić & Divjak, 2018; Israel, 2015). In general, face-to-face learning in a classroom is combined with online learning outside the classroom; participation is typically compulsory for the students. In this setting, teachers support their students on a regular basis and, therefore, dropout rates are low (de Kereki, & Paulos, 2014; de Waard & Demeulenaere, 2017; Filvà, Guerrero, & Forment, 2014; Grella, Staubitz, & Meinel, 2016). Two studies examine specific cases of blended K-12 MOOCs. Grover, Pea, and Cooper (2016) investigate a blended MOOC that is exclusively used inside the classroom. Khalil and Ebner (2015) investigate a MOOC that is embedded in a classroom setting but not compulsory for the students. From the reviewed twenty-four studies, thirteen address blended MOOCs.

On the other hand, there are *non-blended MOOCs* that students can attend in their free time (de Kereki & Manataki, 2016; Hermans

[1] K-12 refers to pre-tertiary education, ranging from kindergarten to the 12th grade.

Table 1
Studies presenting empirical evidence for K-12 MOOCs.

Authors	Blended	Designed for K-12	Topic
Blazquez-Merino et al. (2018)	Yes*	Yes	Physics
de Kereki and Manataki (2016)	No	Yes	CS
de Kereki and Paulos (2014)	Yes	Yes	CS
de Waard and Demeulenaere (2017)+	Yes	No	Foreign languages
Filvà et al. (2014)	Yes	Yes	CS
Grella, Staubitz, Teusner, & Meinel (2016)	No	Yes	CS
Grella, Staubitz, Teusner, & Meinel (2016)	Yes	Yes	CS
Grover et al. (2016)	No**	Yes	CS
Hermans and Aivaloglou (2017)	No	Yes	CS
Janisch et al. (2017)+	Yes	Yes	CS
Khalil and Ebner (2015)	Yes***	Yes	Physics
Koutsakas, Karagiannidis, et al. (2020)	No	Yes	CS
Kurhila and Vihavainen (2015)	No	No	CS
Magen-Nagar and Cohen (2017)	Yes	No	Various sciences
Najafi et al. (2014)	Yes	Yes	Economics
Nigh et al. (2015)	No	No	Pedagogy
Panyajamorn, Kohda, Chongphaisal, & Supnithi (2016)	Yes	No	Chemistry
Perach and Alexandron (2021)+	Yes	No	CS
Politis, Koutsakas, & Karagiannidis (2017)	No	Yes	CS
Sands and Yadav (2020)+	No	No	CS
Staubitz et al. (2019)	Yes	Yes	CS
Tomkins and Getoor (2019)+	No	Yes	CS
Tomkins et al. (2016)	No	Yes	CS
Yin, Adams, Goble, & Vargas-Madriz (2015)	No	Yes	Biology

Note. * blended setting but focus of the study not on the MOOC; ** MOOC used exclusively in the classroom; *** blended setting but participation not compulsory; + Study not included in Koutsakas, Chorozidis, Karamatsouki, and Karagiannidis (2020); CS = Computer Science (e.g., programming, computational thinking, introduction to ICT).

& Aivaloglou, 2017; Politis et al., 2017). Distinguishing between blended and non-blended MOOCs might be important: the learning processes, motivation, and retention of blended and non-blended K-12 MOOCs may be substantially different (de Kereki, & Paulos, 2014; Magen-Nagar & Cohen, 2017). In line with the evidence from the tertiary level, blended K-12 MOOCs could be especially promising.

Seventeen studies investigate MOOCs specifically developed for K-12 education. In general, they aim at secondary level students. The two exceptions are Hermans and Aivaloglou (2017), who compared students younger than twelve with students older than twelve, and Yin, Adams, Goble, & Vargas-Madriz (2015), who relied on a diverse sample of children working on MOOCs together with their parents. The other seven studies investigate MOOCs that do not have a specific target group (de Waard & Demeulenaere, 2017; Panyajamorn et al., 2016; Sands & Yadav, 2020), or that are aimed at university students (Kurhila & Vihavainen, 2015; Magen-Nagar & Cohen, 2017; Perach & Alexandron, 2021) and teachers (Nigh, Pytash, Ferdig, & Merchant, 2015). Since K-12 students are different from adult learners, it may be beneficial to develop specific K-12 MOOCs (Briggs & Crompton, 2016; Filvà et al., 2014; Yin, Adams, Goble, & Vargas-Madriz, 2015).

In terms of topic, sixteen out of the twenty-four studies investigate MOOCs in the realm of computer science, e.g., programming (de Kereki & Manataki, 2016; Grella et al., 2016) or an introduction to the basics (Perach & Alexandron, 2021). Five studies address other STEM topics, such as biology (Yin, Adams, Goble, & Vargas-Madriz, 2015), physics (Blazquez-Merino et al., 2018; Khalil & Ebner, 2015), and chemistry (Panyajamorn et al., 2016). Only three out of the twenty-four studies provide evidence on MOOC learning in the social sciences (de Waard & Demeulenaere, 2017; Najafi, Evans, & Federico, 2014; Nigh et al., 2015).

2.3. Evaluation methods and findings

Concerning qualitative evidence, based on twelve interviews Yin, Adams, Goble, & Vargas-Madriz (2015) report that children use MOOCs differently to adult learners. Children seem to engage less with video lecturers and need more support. Interviews with teachers and students (de Kereki, & Paulos, 2014), and focus groups with students (de Waard & Demeulenaere, 2017), report high perceived and observed learning effects. Grella et al. (2016) conducted a workshop with teachers on K-12 MOOCs, which reveals a high workload for preparing blended MOOC settings. A qualitative analysis of transfer tasks indicates that K-12 MOOC acceptance is higher among more active and motivated students (Nigh et al., 2015).

Concerning quantitative evidence, twelve studies use questionnaires to capture student assessment of K-12 MOOCs. The response rates differ substantially between voluntary (2.7%–9.0%) (de Kereki & Manataki, 2016; Koutsakas, Karagiannidis, et al., 2020) and compulsory questionnaires (100%) (Blazquez-Merino et al., 2018). Overall, students may be willing to learn with MOOCs (de Kereki, & Paulos, 2014; Nigh et al., 2015; Politis et al., 2017); however, evidence for the factors predicting the acceptance of K-12 MOOCs is scarce. An exception is the study by Magen-Nagar and Cohen (2017), who report positive correlations among the students' learning strategies, motivation, and perceived learning effects. However, the authors call for more evidence explaining the learning effects and student motivation in K-12 MOOCs. Moreover, the used MOOC is not deliberately designed for the K-12 level.

A further quantitative evaluation approach is to utilize data from the MOOC platform. Such data is used to investigate retention (de Kereki & Manataki, 2016; Hermans & Aivaloglou, 2017; Khalil & Ebner, 2015; Kurhila & Vihavainen, 2015) and learner activity (Grella et al., 2016; Kurhila & Vihavainen, 2015; Tomkins, Ramesh, & Getoor, 2016; Tomkins & Getoor, 2019). Students who receive support participate more actively and perform better in a posttest (Tomkins et al., 2016; Tomkins & Getoor, 2019). Khalil and Ebner (2015) report that even if a K-12 MOOC is integrated into formal learning (blended), the retention rate is low if participation in the MOOC is voluntary. Students seem to struggle with unsupported MOOC learning (Blazquez-Merino et al., 2018).

Ten studies address learning outcomes in a blended setting and one in a non-blended setting. Five of these eleven studies rely on perceived learning outcomes, i.e., self-assessments (de Kereki, & Paulos, 2014; de Waard & Demeulenaere, 2017; Magen-Nagar & Cohen, 2017; Politis et al., 2017; Staubitz, Teusner, & Meinel, 2019). The results seem promising—students, in general, report substantial perceived learning outcomes. The results from the six studies using objective measurement also report positive outcomes. However, the most extensive study (N = 182) covers a highly specific situation in rural Thailand (Panyajamorn et al., 2016). Blazquez-Merino et al. (2018) compare the performance of (N = 77) students who studied in an online physics lab with (N = 77) students who studied in an offline lab. Students in both groups achieved similar results. However, the MOOC only served to standardize the instruction in order to compare the learning effects of offline and online physics labs. The other four studies are based on rather small samples of K-12 students using a blended MOOC; N = 55 (Perach & Alexandron, 2021), N = 28 (Grover et al., 2016), N = 15 (Najafi et al., 2014), and N = 14 (Janisch, Ebner, & Slany, 2017); the authors call for more research on learning effects. Overall, student support and deliberate integration into the curriculum may be conducive to learning (Najafi et al., 2014; Staubitz et al., 2019).

In summary, the available evidence points to MOOCs as a promising means for K-12 learning, especially in the form of blended MOOCs. However, most studies were carried out in the disciplines of STEM. Studies about student perception of K-12 MOOCs mostly do not rely on established technology acceptance models. Studies that utilize performance tests to assess learning effects often do not even provide enough information to calculate effect sizes, or are based on quite small samples.

2.4. The present study

2.4.1. The i-MOOC

To shed further light on the phenomenon of K-12 MOOCs, the study at hand utilizes the i-MOOC (https://i-mooc.ch/). The i-MOOC aims at fostering Swiss upper secondary level students' information literacy (IL). IL is a learning goal for these students specified in the relevant curricula (BBT, 2006; EDK, 1994; SBFI, 2013). To create awareness for the i-MOOC, we contacted all pertinent schools in German-speaking Switzerland. Interested teachers contacted the project team and got access to the i-MOOC free of charge. Moreover, teachers received technical support and pedagogical advice upon request.

The i-MOOC is a blended MOOC. The students were enrolled and supported by their teachers during in-class instruction. The theoretical background of the i-MOOC is the 7i framework (Seufert, Guggemos, Moser, & Sonderegger, 2019). It specifies what constitutes IL and the learning goals that should be pursued. The i-MOOC covers these learning goals by means of six modules. The intended time to fully complete the i-MOOC is 6 h. However, depending on the level of support, some teachers used up to 7.5 h.

Information literacy may be a suitable content for a K-12 MOOC: IL is regarded as a key 21st-century skill and an important educational goal (Fraillon, Ainley, Schulz, Duckworth, & Friedman, 2019; Oberländer, Beinicke, & Bipp, 2020; Yang et al., 2021). Indeed, it may be important to foster students' information literacy; the idea of digital natives who competently use technology without instruction is a myth (Kirschner & Bruyckere, 2017).

The i-MOOC was implemented using the OpenEdX platform and hosted by Swissuniversities. From a technical and conceptual point of view, the i-MOOC could be taken by a virtually unlimited number of students. It comprises explanatory videos, information material, cheat sheets, quizzes, and peer feedback tasks. For MOOCs, in particular, it is important to ensure a high-quality course design (Kim et al., 2021). In light of this, we developed the i-MOOC using a design research approach. Based on criteria for high-quality MOOCs (Egloffstein, Koegler, & Ifenthaler, 2019; Margaryan, Bianco, & Littlejohn, 2015), we developed, evaluated, and refined the initial version. For a detailed description of the design research process of developing and improving the i-MOOC, see Seufert, Guggemos, Moser, & Sonderegger (2019) and Moser, Guggemos, & Seufert (2021). Supplementary material S2 outlines how the design criteria for high-quality MOOCs are considered in the i-MOOC.

2.5. Theoretical framework for evaluation

According to Whetten (1989), the crucial point in theory development is a good trade-off between comprehensiveness and parsimony. This trade-off guided our evaluation strategy. Overall, we aim at answering three related research questions:

RQ1: What is the learning effect of a K-12 MOOC?

RQ2: What are the predictors of student acceptance of a K-12 MOOC?

RQ3: What are the predictors of the learning effect of a K-12 MOOC?

Concerning MOOC research, Reich (2015) calls for studies that put learning outcomes into focus. A generally applied approach to evaluate educational technology interventions is to determine the achieved effect size of this intervention (Scheiter, 2021). An effect size can be based on the comparison of student posttest scores with their pretest scores, or comparing the scores of a treatment group with those of a control group after the intervention (Hattie, 2009, p. 8). To claim effectiveness, it may not be sufficient to demonstrate a statistically significant positive effect. Rather, an effect size of greater than 0.4 could be in the desirable zone (Hattie, 2009; Reeves & Lin, 2020). Since the i-MOOC was thoroughly developed based on established design principles, we hypothesize an effect size (Cohen's d) in the zone of desirable effects, i.e., statistically significantly greater than 0.4.

To evaluate educational technology, investigating student technology acceptance is an established approach (Al-Emran, Mezhuyev, & Kamaludin, 2018; McGill, Klobas, & Renzi, 2014; Raffaghelli, Rodríguez, Guerrero-Roldán, & Bañeres, 2022). However, for blended K-12 MOOCs the nature of technology acceptance may be different from general technology. In this case, teachers decide to use, or not to use, a MOOC in their instruction. The actual use is not within the control of the students. Hence, the acceptance of K-12 MOOCs could address whether students have the intention to use, or want their teachers to use, MOOCs. Although students do not decide on the use, a lack of student technology acceptance could hinder the adoption of K-12 MOOCs. Teachers, whose students are reluctant to study with MOOCs, may not use them (in the future). To evaluate student technology acceptance, we rely on the unified theory of acceptance and use of technology (UTAUT) (Venkatesh, Morris, Davis, & Davis, 2003). The UTAUT is a well-founded theory suitable for educational technology research (Novak, 2021; Valtonen et al., 2022); it integrates other theories such as the technology acceptance model. The UTAUT implies that student behavioral intention, e.g., to use a MOOC (technology acceptance), can be explained by performance expectancy, effort expectancy, and social influence. The UTAUT has been extended to the UTAUT2 (Venkatesh, Thong, Xu, Venkatesh, & Xu, 2012). UTAUT2 constructs, definitions, and items are reported in Table 2. From the additional predictors implied by the UTAUT2, beyond the UTAUT, only hedonic motivation may be relevant in our context. We do not consider facilitating conditions because all necessary resources for using the i-MOOC, such as digital devices, are available to the students. In Switzerland, 99% of upper secondary level students own a smartphone (Bernath et al., 2020) and this allows them to use the i-MOOC anywhere. In our case, habits do not play a role as blended K-12 MOOCs are a new phenomenon for the students. Price value is irrelevant because students do not have to pay to use the i-MOOC. Overall, the hypothesized positive predictors for student acceptance of K-12 MOOCs are performance expectancy, effort expectancy, social influence, and hedonic motivation.

The UTAUT is, as the name implies, a unified theory that integrates several theories (Novak, 2021). It is informed by motivation theories (Venkatesh et al., 2003, 2012). The expectancy-value theory is among the most established motivation theories (Wigfield, Muenks, & Eccles, 2021). It implies that performance can be predicted by expectancies for success and the task values. Ranellucci, Rosenberg, and Poitras (2020) argue that although technology acceptance and expectancy-value literature both use different terminology, they may be compatible in nature. The predictors of technology acceptance implied by the UTAUT2 could be regarded as manifestations of either expectancies of success or task values. In sum, we hypothesize performance expectancy, effort expectancy, social influence, and hedonic motivation to positively predict IL gain. Relying on these technology acceptance predictors may allow for a good trade-off between comprehensiveness and parsimony of the overall evaluation strategy.

The UTAUT2 proposes three moderators of the relationships between the constructs: age, gender, and experience. In the specific circumstances of our study, technology experience can be excluded because the phenomenon blended MOOC is new to the students in the sample. However, we add school type as a potential moderator; school type is regularly considered in meta-analyses (e.g., concerning the association between motivation and performance, see Kriegbaum, Becker, & Spinath, 2018).

3. Method

3.1. Sample

In 2020, 1032 students, nested in fifty-six classes, attended the i-MOOC; 880 students completed the IL pretest and posttest integrated into the i-MOOC. All 880 students were subsequently asked to fill in a questionnaire; participation was entirely voluntarily and follows the ethical standards of the University of St.Gallen. Overall, 170 students filled in the questionnaire. The response rate equals 19.3%, which is higher than in all the reviewed K-12 MOOCs where filling in the questionnaire was also voluntary (max = 9.0%). By means of Mahalanobis distances, we checked for outliers (Leys, Delacre, Mora, Lakens, & Ley, 2019). We identified three statistically significant ($\alpha = 0.01$) outliers that we inspected. These three students, for instance, did not show any variance in their answers. We concluded there was a lack of sufficient test motivation for these students and, consequentially, we excluded all three of them. The final sample comprises N = 167 upper secondary level students nested in thirty-one classes. This implies a substantial reduction in our sample (from 1032 to 167); we will come back to this point in section 5. From these 167 students, 1.3% of the data is missing. Based on Little's test of missing completely at random ($\chi^2(80) = 75.3$, p = .628), as well as an inspection of the data, we might conclude this data is missing completely at random. The average age of the students was 17.80 years (SD = 2.19), median = 17 years, mode = 16 years, min = 16 years, max = 25 years; 51% of the students were female; 49% attended a vocational school, 44% a high school, and 7% other school types.

3.2. Data analysis

The IL pretest and posttest scores are the basis for answering RQ1 and RQ3. Considering these as latent variables in subsequent analyses is not possible due to our rather small sample size. However, we relied on factor scores as they are superior to sum scores (McNeish & Wolf, 2020). Furthermore, using factor scores based on confirmatory factor analysis allows us to assess the overall fit of the measurement model. We can also demonstrate measurement invariance over time, which is necessary to allow for a valid comparison of pre- and posttest scores (van de Schoot, Lugtig, & Hox, 2012). To check for measurement invariance over time, we use the approach of van de Schoot et al. (2012). We compare an unrestricted model, where all parameters for the pre- and posttest are individually

Table 2
Constructs and items for evaluating technology acceptance based on the UTAUT(2) (Venkatesh et al., 2003, 2012).

Construct	Definition	no.	Item	M	SD	Min	Max
Performance expectancy	Degree to which MOOCs are perceived as valuable for learning	1	The i-MOOC helps me to learn more.	3.3	1.6	1	7
		2	The i-MOOC helps me to learn faster.	3.3	1.6	1	7
		3	The i-MOOC is useful for learning.	3.7	1.8	1	7
Effort expectancy	Degree of ease associated with the use of MOOCs	4	I can use the i-MOOC without help from others.	5.4	1.8	1	7
		5	It is easy for me to use the i-MOOC.	3.5	1.7	1	7
Social influence	Degree to which students believe significant others want them to use MOOCs	6	People I care about think I should study with online courses, like the i-MOOC.	2.5	1.7	1	7
		7	My parents think I should study with online courses, like the i-MOOC.	2.4	1.7	1	7
		8	My friends think I should study with online courses, like the i-MOOC.	2.3	1.6	1	7
Hedonic motivation	Degree of fun or pleasure derived from studying with MOOCs	9	Studying with the i-MOOC is fun.	2.9	1.8	1	7
		10	Studying with online courses, like the i-MOOC, would make school more entertaining.	3.1	1.9	1	7
Technology acceptance	Degree of intended use of MOOCs or appreciation of their use by teachers	11	If possible, I would study more with online courses, like the i-MOOC.	2.9	1.9	1	7
		12	I would like my teachers to use more online courses, like the i-MOOC.	2.9	1.9	1	7

Note. The items were presented in random order to the students and measured on a seven-point scale of rating, ranging from entire disagreement to entire agreement.

estimated, with a model where intercepts and loadings are restricted to be equal across both tests. A significant χ^2-test would indicate a violation of scalar measurement invariance. In this case, we free parameters across groups to achieve partial measurement invariance; this could allow for a valid comparison across the two timepoints (van de Schoot et al., 2012). We use a robust maximum likelihood (MLR) estimator in the confirmatory factor analysis because the data are (by nature) not normally distributed.[2] Moreover, this allows us to handle missing data (test items not answered) by means of full information maximum likelihood (Jia & Wu, 2019). This approach is superior to simply handling missing data as incorrect answers (Köhler, Pohl, & Carstensen, 2017). We perform all these analyses using the R package lavaan 0.6–9 (Rosseel, 2012).

To answer RQ1, we rely on paired t-tests to assess the statistical significance, and Cohen's d to quantify the effect size of the increase in IL. Since the IL test scores are not normally distributed (Shapiro-Wilk test: $p < .001$) and clustered (students nested in classes), we use bootstrapping to calculate standard errors using the R package MKinfer 0.6 (Kohl, 2020). We opt for using Bartlett factor scores as they yield unbiased estimators for the true scores (DiStefano, Zhu, & Mîndrilã, 2009).

Concerning RQ2 and RQ3, we use covariance-based structural equation modeling to predict acceptance of K-12 MOOCs, and the IL posttest score (R package lavaan 0.6–9). Missing data are handled with full information maximum likelihood. To assess the quality of the measurement models, we rely on the cut-off values proposed by Hair, Risher, Sarstedt, and Ringle (2019), and for the overall model-fit, on the cut-off values of Hu and Bentler (1999). Since our data is clustered, we use cluster-robust standard errors (McNeish, Stapleton, & Silverman, 2017). Moreover, teachers can use the i-MOOC in various ways. In particular, they can offer their classes different degrees of guidance and support. Since this between-class heterogeneity is likely to influence the perception of the i-MOOC, we use group-mean centered variables in our structural equation models (McNeish & Kelley, 2019), i.e., for answering RQ2 and RQ3. This means every student is compared with their classmates. Between-class heterogeneity is removed and we can analyze pure individual-level effects. This approach may be suitable, in particular, for small samples (McNeish & Kelley, 2019). Class-mean centering, indeed, seems to be necessary as the between-class variance (ICC1) is substantial for the posttest score of IL: 19.3%. For the pretest score, the ICC1 equals 7.3%. For the UTAUT2 variables, it ranges between 0.0% (item 10, see Table 2) and 12.1% (item 5, see Table 2) with a mean of 4.8% ($SD = 4.0$). To control for moderators implied by the UTAUT2, we carry out multigroup analysis. The groups are: gender, male vs. female; age, $\geq$18 years vs. $<$18 years; and school type, high schools vs. vocational schools.

Concerning RQ3, we use a regressor-variable model (Aichele, Hartig, & Michaelis, 2021), where the IL posttest score is used as the dependent variable, and the pretest score, as well as the UTAUT2 predictors, as independent variables. Following Skrondal and Laake (2001), we use regression factor scores for the independent variable (IL pretest score) and the Bartlett factor score for the dependent variable (IL posttest score). All other variables are latent.

3.3. Measurement instruments

To measure student IL, both before and after the i-MOOC, we utilize a short version of a validated performance test (Seufert, Stanoevska-Slabeva, & Guggemos, 2020) that covers the 7i framework of IL. The items can be found in the supplementary material S3. It comprises twenty-two binary items. Using a performance test instead of self-assessment may be important because students' ability to correctly evaluate their IL seems to be low (Mahmood, 2016). Since both the i-MOOC and the IL test are based on the 7i framework, we may have achieved instructional sensitivity (Deutscher & Winther, 2018; Naumann, Rieser, Musow, Hochweber, & Hartig, 2019):

[2] A 'weighted least square mean and variance adjusted' (WLSMV) estimator (C.-H. Li, 2016) yields identical ($r = 0.997$) results.

Sample item of the information literacy performance test.

In the afternoon, your best friend calls and asks you: "Can you please help me with my research on the Swiss President? I have to give a 15-minute lecture on him tomorrow in class."
What is the most important thing you need to find out to help your friend with his problem?

- Where I can find information about the Swiss President
- What my friend has already found out
- Whether the Swiss President has a profile on a social media platform (e.g., Instagram, Twitter)
- What exactly the assignment is

Corresponding item from the formative quiz in the i-MOOC.

It is important to identify my information needs because ...

- I can search for information goal-oriented.
- I like it.
- it allows to evaluate the usefulness of found information.
- it helps to identify fake-news.

Corresponding instruction material in the i-MOOC.

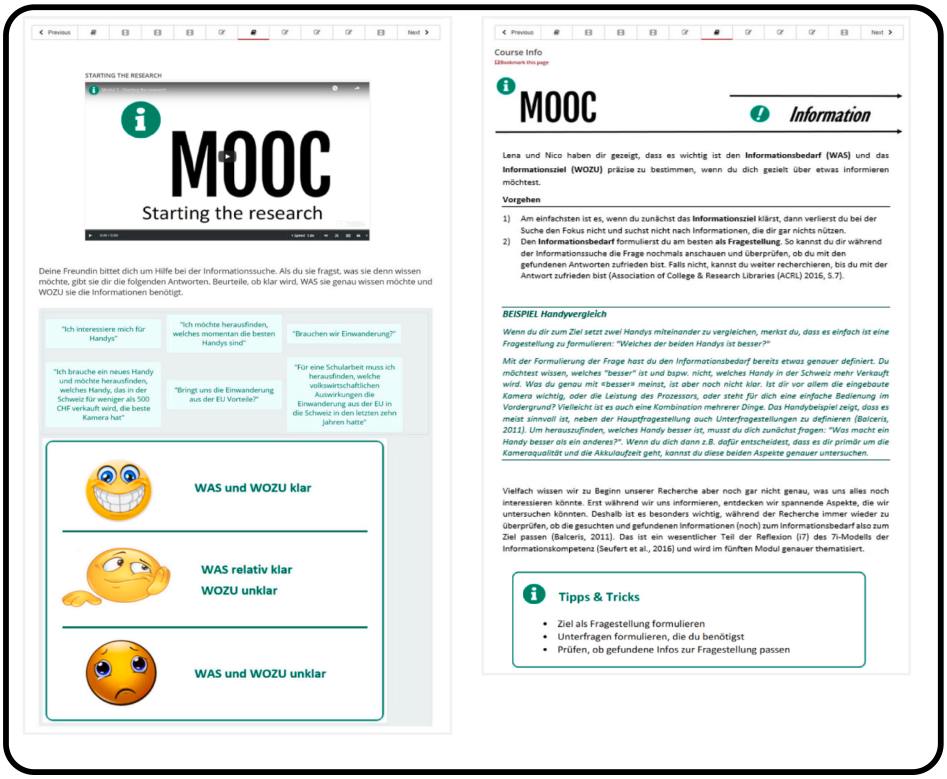


Fig. 1. Sample item of the information literacy performance test, corresponding quiz item, and instruction material in the i-MOOC.

the 7i framework acts as the curriculum, i.e., it specifies what kind of task students should be able to carry out in order to demonstrate IL. The i-MOOC is designed to enable students to master these tasks, which can be assessed with the used IL test. On the other hand, however, teaching to the test has to be avoided (Popham, 2001). In our context, teaching to the test would imply that course material and quizzes are similar to the test items. To prevent this, the IL performance test and the quizzes in the i-MOOC are based on different content. An example of an item from the IL test, as well as the corresponding instruction and quiz item from the i-MOOC, can be found in Fig. 1. In addition, students did not receive feedback whether answers to specific items in the pretest were correct in order to avoid learning effects due to test taking. However, they did receive an overall percentage score to gauge their level of IL before starting the i-MOOC.

We adapted the UTAUT and UTAUT2 questionnaires (Venkatesh et al., 2003, 2012) to our context. The questions are depicted in Table 2. They are measured on seven-point scales of rating, ranging from entire disagreement to entire agreement.

4. Results

4.1. Assessment of measurement instruments

The joint confirmatory factor analysis of the IL pretest and posttest yielded a decent fit: YB-χ^2 (392) = 398.955, p = .393, CFI = 0.987, TLI = 0.985, RMSEA = 0.010, SRMR = 0.057. However, scalar measurement invariance (equal intercepts and loadings) across time is not achieved: χ^2 (42) = 62.163, p = .023. After allowing three intercepts of items to be different across the pre- and posttest, the difference in fit between the unrestricted model and the restricted model, ensuring partial measurement invariance, is no longer significant: χ^2 (39) = 30.978, p = .820.

Apart from two exceptions, the measurement models are sound (see Table 3). Internal consistency reliability is decent with α and ω greater than 0.7. Convergent validity is achieved as the average variance extracted is greater than 0.5. Discriminant validity might also be ensured; the highest heterotrait–monotrait ratio equals 0.74 (between performance expectancy and technology acceptance). This is well below the cut-off value of 0.9 for conceptually similar constructs. The two exceptions are: first, the internal consistency reliability of the pretest IL score (0.67) is slightly below the cut-off value of 0.7; second, there is a lack of discriminant validity between hedonic motivation and the acceptance of K-12 MOOCs. This can be seen from a heterotrait–monotrait ratio of 0.97, 95% CI [0.94, 1.01], i.e., not significantly smaller than 1. This means these two constructs, albeit different from a conceptual point of view (UTAUT2), are hard to empirically separate. We will come back to this point in the discussion section.

4.2. Learning effect of information literacy K-12 MOOC (RQ1)

On a descriptive level, in the pretest the students scored on average 10.38 out of 22 points (SD = 3.51), median = 11 points, min = 1 point, max = 18 points. In the posttest, they scored on average 12.89 (SD = 3.99) points, median = 14 points, min = 1 point, max = 22 points. This equals a medium effect: d = 0.66, 95% CI [0.43, 0.89], $p < .001$. Based on Bartlett factor scores and established partial measurement invariance, the posttest score is significantly higher than the pretest score with a medium effect size: d = 0.75, 95% CI [0.53, 0.97], $p < .001$. As the confidence interval does not include 0.4, the effect is significantly higher than 0.4 and, hence, in the zone of desirable effects.

There are no differences concerning age, gender, and school type in terms of the increase in IL ($p > .352$). Fig. 2 shows the normalized learning gain in IL grouped by gender and type of school based on Bartlett factor scores. The normalized learning gain (g) is defined as: g = (post – pre)/(100 – pre), where post- and pretest scores are given as a percentage (Marx & Cummings, 2007). The median normalized learning gain across all groups equals 24.9%, depicted as a horizontal line in Fig. 2.

Table 3
Assessment of the measurement models (N = 167).

	Construct	Sum/Mean	SD	α	ω	AVE	Below diagonal correlations; above diagonal HTMT						
							1	2	3	4	5	6	7
1	Pretest score IL	10.4	3.5	.64	.67	–	1	–	–	–	–	–	–
2	Posttest score IL	12.9	4.0	.75	.78	–	**.50**	1	–	–	–	–	–
3	Performance expectany	3.4	1.5	.90	.90	.75	-.06	**.21**	1	.45	.56	.74	.64
4	Effort expectancy	4.9	1.6	.74	.76	.62	.01	**.29**	**.44**	1	.24	.35	.28
5	Social influence	2.4	1.5	.91	.91	.77	.04	**.20**	**.56**	**.26**	1	.64	.57
6	Hedonic motivation	3.0	1.7	.87	.87	.92	-.04	.08	**.75**	**.37**	**.65**	1	.97
7	Technology acceptance	2.9	1.8	.95	.95	.77	-.06	.07	**.63**	**.29**	**.56**	**.97**	1

Note. Sum/Mean = sum of correctly answered binary test items out of 22/mean of UTAUT2 variables (see Table 2), SD = standard deviation, α = Cronbach's alpha, ω = McDonald's omega total, AVE = average variance extracted, HTMT = heterotrait-monotrait ratio. The correlations are based on group-mean centered variables (group = class). All correlation in bold are significant at the 5% level (two-sided) considering the cluster structure of the data (students nested in classes). Pretest scores are regression factor scores, posttest scores are Bartlett factor scores; the UTAUT2 constructs 3–7 are latent variables measured on a seven-point scale of rating.

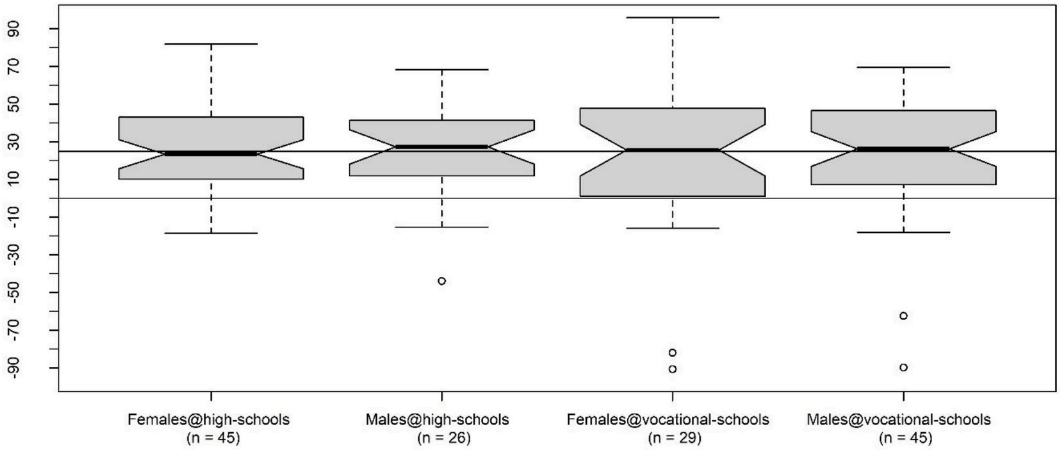


Fig. 2. Information literacy normalized learning gain grouped by gender and school type.

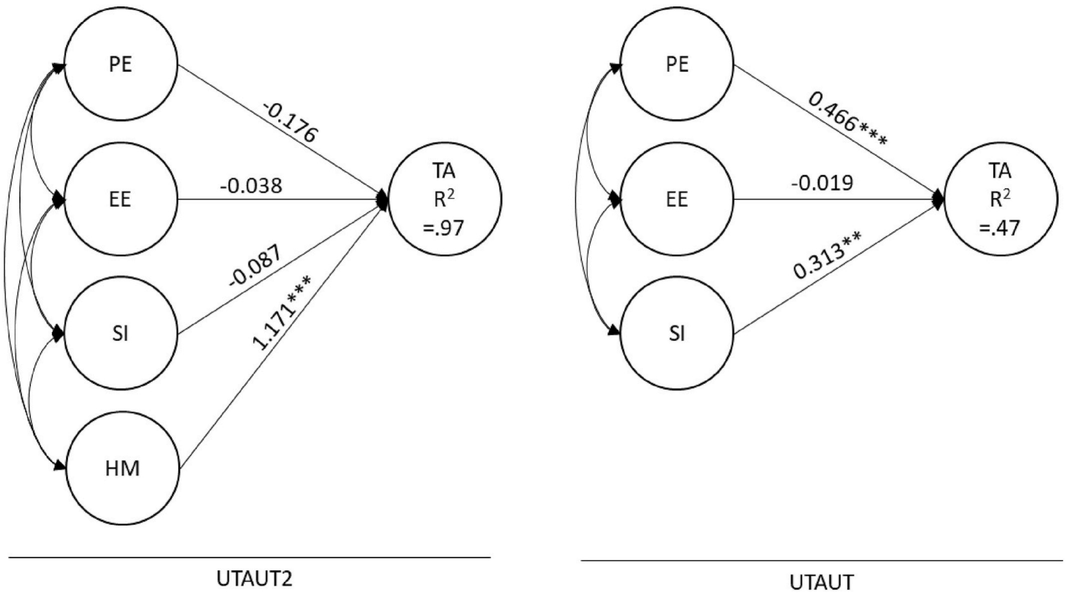


Fig. 3. Structural equation models to predict student acceptance of K-12 MOOCs (N = 167). PE = Performance Expectancy, SI = Social Influence, EE = Effort Expectancy, HE = Hedonic Motivation, TA = Technology Acceptance. Note. Standardized path coefficients based on group-mean centered-variables. Assessment of significance based on cluster robust standard errors (cluster = classes). * $p < .05$, ** $p < .01$, *** $p < .001$. Fit values of UTAUT2 model: YB-χ^2 (44) = 52.38 ($p = .181$), CFI = 0.995, TLI = 0.992, RMSEA = 0.034 (90% CI [0.000, 0.065]), SRMR = 0.026. Fit values of UTAUT model: YB-χ^2 (29) = 30.08 ($p = .410$), CFI = 0.999, TLI = 0.999, RMSEA = 0.015 (90% CI [0.000, 0.064]), SRMR = 0.023. All factor loadings are significant ($p < .001$).

4.3. Predicting acceptance of K-12 MOOC (RQ2)

Student acceptance of our K-12 MOOC is rather low. Based on a seven-point scale of rating, it equals 2.89 (SD = 1.82), median = 2.50, min = 1, max = 7. The fit values of the UTAUT2 model are excellent: CFI = 0.995, TLI = 0.992, RMSEA = 0.034, SRMR = 0.026. Fig. 3 depicts this model. As can be seen, the only significant predictor for student acceptance of K-12 MOOCs is hedonic motivation ($\beta = 1.171$, $p < .001$); the explained variance equals 97%. Due to the identified lack of discriminant validity, we also estimated the UTAUT model, which does not consider hedonic motivation (see Fig. 3). In this model, performance expectancy ($\beta = 0.466$, $p < .001$) and social influence ($\beta = 0.313$, $p = .002$) significantly predict student acceptance of K-12 MOOCs. However, effort expectancy is not a significant predictor ($\beta = -0.019$, $p = .847$).

The multigroup analyses, considering measurement invariance, did not yield overall significant differences: gender ($\Delta\chi^2$ (58) = 71.638, $p = .108$), age ($\Delta\chi^2$ (58) = 60.049, $p = .401$), and school type ($\Delta\chi^2$ (58) = 61.419, $p = .355$).

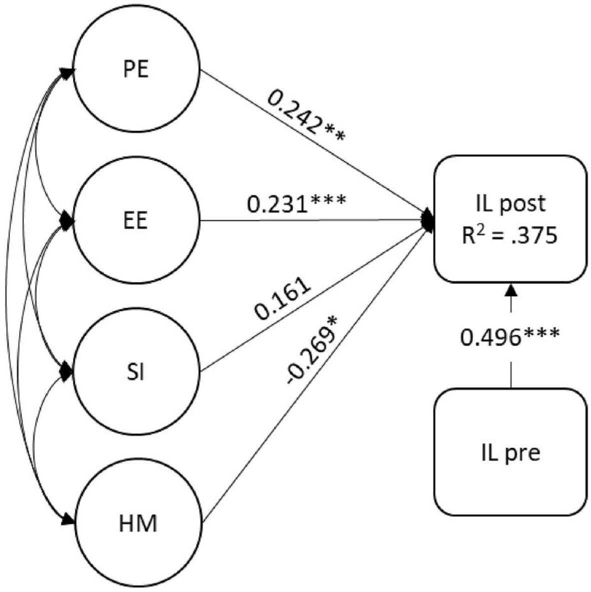


Fig. 4. Structural equation model to predict Information Literacy (IL) after attending the i-MOOC (N = 167). IL post = IL posttest score (Bartlett factor score), IL pre = IL pretest score (regression factor score), PE = Performance Expectancy, EE = Effort Expectancy, SI = Social Influence, HM = Hedonic Motivation. Note. Standardized path coefficients. Assessment of significance based on cluster robust standard errors (cluster = classes). * $p < .05$, ** $p < .01$, *** $p < .001$. Fit values: YB-χ^2 (41) = 53.52 (p = .091), CFI = 0.991, TLI = 0.985, RMSEA = 0.043 (90% CI [0.000, 0.067]), SRMR = 0.028. All factor loadings are significant ($p < .001$).

4.4. Predicting information literacy after K-12 MOOC attendance (RQ3)

Fig. 4 shows the model used to predict student performance in the IL posttest when controlling for performance in the pretest. The fit is excellent: CFI = 0.991, TLI = 0.985, RMSEA = 0.043, SRMR = 0.028. Performance expectancy (β = 0.242, p = .009) and effort expectancy (β = 0.231, $p < .001$) positively predict the IL posttest score when controlling for the pretest score. However, hedonic motivation is a negative predictor: β = −0.269, p = .034.

The multigroup analyses, considering measurement invariance, did not yield overall significant differences: gender ($\Delta\chi^2$ (53) = 62.485, p = .175), age ($\Delta\chi^2$ (53) = 61.530, p = .197), and school type ($\Delta\chi^2$ (53) = 52.250, p = .503).

5. Post hoc power analysis and robustness checks

Our sample may be regarded as small ($N < 200$). However, the required sample size depends on the specific situation (Wolf, Harrington, Clark, & Miller, 2013). Hence, we carried out a post hoc power analysis. Concerning RQ1, we can detect a small effect (d = 0.21) with a probability of 80%. To determine the achieved power for answering RQ2 and RQ3, we used the powerSEM 1.1.0 (Moshagen & Erdfelder, 2016) and pwrSEM 0.1.2 (Y. A. Wang & Rhemtulla, 2021) packages in R. We can identify a misspecified model (RMSEA > 0.08) with a probability of 97% and 93%, respectively. Assuming factor loadings of 0.7 for all the UTAUT2 constructs, which implies a sufficient convergent validity of the measurement model (Hair et al., 2019), we can detect standardized path coefficients of 0.300 with a probability of at least 86% and 73%, respectively. Standardized path coefficients of greater than 0.300 may be reasonable (Scherer, Siddiq, & Tondeur, 2019). Overall, our sample may yield sufficient power to answer our research questions.

Concerning RQ3, we used a regressor-variable model, i.e., the pretest score is used as an independent variable to predict the posttest score (see Fig. 4). The alternative specification in a pre-posttest design would be a change-score model, where the difference between posttest and pretest scores acts as the dependent variable. Both models have specific assumptions; it might be beneficial to estimate both models and check if they agree (van Breukelen, 2013). In our case, the results are identical in terms of sign and statistical significance. The estimated associations with the change score are: performance expectancy (β = 0.285, p = .009), effort expectancy (β = 0.265, $p < .001$), social influence (β = 0.173, p = .179), and hedonic motivation (β = −0.301, p = .035). As the change score is the dependent variable in the change score model, we used Bartlett factor scores to form the change score (Skrondal & Laake, 2001). Since the two models yield identical results, two claims may be valid: 1) performance expectancy predicts the IL posttest score while controlling for the pretest score, and 2) performance expectancy predicts the IL gain or learning effect. Due to its more vivid interpretation, we will stick to the latter claim in the discussion section.

Although superior to sum scores, factor scores are not free of measurement errors (McNeish & Wolf, 2020). Hence, we used the approach suggested by Savalei (2019) to check whether considering measurement error has an impact on our findings. We restricted the residual variances of the pretest and posttest scores to values corresponding with an internal consistency reliability of 0.8 (reasonable maximum). The findings are identical in terms of sign and statistical significance, e.g., technology acceptance on hedonic motivation: $\beta = -0.294$, $p = .033$.

Our sample may suffer from a self-selection bias. Hence, we checked to see if the 167 students in our sample are different from those students who did not fill in the questionnaire. There are no significant differences in terms of gender (χ^2 (1) = 3.646, p = .056), age group (χ^2 (1) = 1.441, p = .230), and school type (χ^2 (1) = 0.032, p = .856) between the two groups. Although not statistically significant at the 5% level, there is a tendency for females to volunteer more often to fill in the questionnaire. The finding of non-significant differences may be important because these variables could moderate the reported relationships when answering RQ2 and RQ3 (Kriegbaum et al., 2018; Venkatesh et al., 2012). Besides this, we compared the pretest and posttest scores of the students who filled in the questionnaire with those who refused to do so. T-tests with bootstrap standard errors show no differences concerning pretest score ($d = 0.00$, $p = .995$) and posttest score ($d = -0.09$, $p = .464$). Hence, the answer to RQ1 may not be influenced by student self-selection.

6. Discussion

6.1. Findings

Concerning RQ1, the learning effect is $d = 0.75$, which is in the zone of desirable effects (Hattie, 2009; Reeves & Lin, 2020). However, other K-12 MOOCs yield even higher effect sizes based on a pretest-posttest comparison ($d = 2.4$; Grover et al., 2016). On a descriptive level, the learning gain may be regarded as modest: on average, the students scored 10.38 points in the pretest and 12.89 points in the posttest, out of 22 points. However, comparisons based on descriptive statistics may be misleading because they do not consider measurement invariance and are based on sum scores. Nevertheless, there may be reasons on the content level for our findings. The quality of educational technology use and the expected learning gains depend on teachers' technology-related knowledge, skills, and attitudes (Scheiter, 2021; Seufert, Guggemos, & Sailer, 2021). K-12 MOOCs are quite a new phenomenon for teachers. Education and training on how to use blended MOOCs in an effective way may be necessary to achieve high levels of student learning. In general, students' IL seems to be hard to improve. Although there may be consent about the importance of fostering IL among students, and despite the fact IL is part of national curricula, it does not seem to increase markedly over time (Fraillon et al., 2019).

The learning effect does not significantly differ among gender, school type, and age groups. By nature (being massive), MOOCs are designed for a broad target group. In the case of the i-MOOC, the target group consists of upper secondary level students in German-speaking countries. Although students in our sample are from school types that are quite different in nature (vocational schools and high schools), the i-MOOC yielded similar learning gains. Hence, it may be possible to develop a MOOC that suits a broad variety of upper secondary level students.

The effect size of $d = 0.75$ is based on a pretest-posttest comparison, i.e., there is no control group. Students' IL may also increase without attending the i-MOOC. However, Seufert, Stanoevska-Slabeva, & Guggemos (2020) reported that one school year yields an IL increase of $d = 0.1$ ($p < .01$). Against this backdrop, the i-MOOC may yield learning effects that are greater than the effects of general IL instruction during the course of one school year. Nevertheless, we cannot answer how the i-MOOC compares to other instructional means. For instance, teachers could use the 6 h of time that is required for the i-MOOC for unplugged IL instruction activities. Blazquez-Merino et al. (2018), however, reported similar learning effects of a K-12 MOOC in comparison to unplugged learning activities. Besides this, when evaluating an effect size, it is important to consider cost (Hattie, 2009). A K-12 MOOC may yield smaller learning effects than the well-developed instructional design of a teacher for a specific class; however, the MOOC can, with little effort, be used by a massive number of teachers and, hence, students.

Concerning RQ2, when predicting technology acceptance, there is an issue with discriminant validity in the UTAUT2 model. Adjustments to hypothesized models should be made in line with theory (Henseler, Ringle, & Sarstedt, 2015). Hence, we also estimated the UTAUT model by removing hedonic motivation from the UTAUT2 model. However, the finding of a very high correlation between hedonic motivation and technology acceptance may also be revealing. Since the constructs are conceptually different, as can be concluded from the items listed in Table 2, student acceptance of K-12 MOOCs may be purely driven by hedonic motivation. When this type of motivation is removed (UTAUT), the three remaining predictors—performance expectancy, effort expectancy, and social influence —can explain 47% of the technology acceptance variance. This is in line with technology acceptance studies in the context of education (Scherer et al., 2019). The best predictor in the UTAUT model is performance expectancy, i.e., the perception of MOOCs as a valuable learning resource. However, performance expectancy is on a rather low level ($M = 3.40$); the neutral scale mean equals 4. The perception of significant others (social influence) is also a notable predictor and on a low level ($M = 2.41$). Effort expectancy is the only construct with an overall positive evaluation ($M = 4.90$). However, other than hypothesized, effort expectancy is not a significant predictor for technology acceptance. Overall, student technology acceptance is below the neutral scale mean ($M = 2.89$). In contrast to this, all our reviewed studies report a favorable acceptance of K-12 MOOCs. However, these studies rely on a non-blended (voluntary) MOOC (Nigh et al., 2015), a blended MOOC with students from an elective course (Janisch et al., 2017), and a specific setting in rural Thailand (Panyajamorn et al., 2016). Our sample with students from thirty-one classes may provide a more realistic picture of student acceptance of K-12 MOOCs.

Concerning RQ3, as hypothesized, performance expectancy is a positive predictor for IL gain. Moreover, effort expectancy is a positive predictor for IL gain. This may also be in line with the cognitive load theory (Sweller, 2020): if a MOOC were difficult to use, this could imply extraneous cognitive load that hinders learning. Hedonic motivation, however, is not a positive predictor for IL gain, but negatively predicts it. This may be, at first glance, contradictory to findings in the context of academic performance where hedonic motivation predicts learning gain (Retelsdorf, Köller, & Möller, 2011). However, hedonic motivation in these cases addresses the learning content, e.g., reading. In our case, hedonic motivation addresses the means of instruction, namely, the i-MOOC. Our finding may be in line with Kirschner who argues that it is a myth that learning has to be fun to be efficient; rather, learning has to be rewarding (Kirschner, Verschaffel, Star, & van Dooren, 2017). Hedonic motivation does not predict IL gain when considered as the sole predictor. However, when considering performance expectancy, it negatively predicts it. Performance expectancy and hedonic motivation overlap ($\rho = 0.75$, $p < .001$). Pure hedonic motivation may be detrimental to learning whereas if a K-12 MOOC is perceived as rewarding (performance expectancy), this is conducive to learning. Indeed, studying with the i-MOOC may be hard work. Students have to carry out several (peer-feedback) tasks and answer quizzes. Students who overly assess the i-MOOC as fun and entertaining may be likely to only use it on a superficial level.

6.2. Limitations and recommendations for further research

Our study is not without limitations. We discuss them following Cachero, Barra, Melia, and Lopez (2020), by addressing internal, external, construct, and conclusion validity.

Compared to randomized field experiments as the gold standard, internal validity is suboptimal. We cannot claim that attendance of the i-MOOC caused the increase in IL as we do not have a control group (RQ1). Concerning technology acceptance (RQ2), data are cross-sectional, which does not allow for causal inference. Concerning the predictors for the posttest score (RQ3), there are further ones that may be considered. Such a predictor could be intrinsic value, which goes beyond hedonic motivation (Khechine, Raymond, & Augier, 2020). Again, the reported associations are correlational.

We collected data in a low stakes test environment. In particular, decreasing test motivation could downwardly bias effects (Finney, Sundre, Swain, & Williams, 2016). Fig. 2 may indicate that there are students with a substantial decrease in test motivation. Although we were able to demonstrate that students in our final sample are not different from those who refrained to fill in the questionnaire, in terms of observable characteristics, we cannot rule out differences in terms of unobservable characteristics.

A threat to external validity could be our sample. It is narrow in scope as it comprises only students from German-speaking Switzerland and only one K-12 MOOC. This sample is not representative of upper secondary level students in general. To make general claims about these students and K-12 MOOCs, replication studies would be necessary. Moreover, if MOOCs became a common phenomenon in K-12 education, it would be necessary to consider the level of habit as a predictor for technology acceptance; facilitating conditions should be considered if access to digital devices is limited (Venkatesh et al., 2012).

Concerning construct validity, a limitation is our reliance on relatively short IL tests and technology-acceptance scales. A reason for this is that in the first design research circle, teachers strongly requested a reduction in the quantity of research-related questions. Moreover, we only considered direct learning effects. However, there are likely to be positive side effects of using a MOOC, e.g., fostering students' digital skills. Further research could aim at a comprehensive picture of the learning outcomes of a K-12 MOOC.

Conclusion validity may be impaired as we only investigated individual-level associations, i.e., the student level. It is likely that class-level variables moderate the associations at the individual level. A likely moderator is the level of guidance and support offered by the teacher in a specific class. We were not able to model this multilevel structure due to our small class-level sample size ($N = 31$) (McNeish & Stapleton, 2016). Nevertheless, our (statistical) inferences are valid because we relied on class-mean centered variables and cluster-robust standard errors (McNeish et al., 2017; McNeish & Kelley, 2019). This approach may be the most suitable one for our data structure. However, for a comprehensive understanding of technology integration, it would be important to also consider teacher variables in the quantitative design. The effectiveness of educational technology use is likely to be moderated by the technology-related knowledge, skills, and attitudes of teachers (Scheiter, 2021; Seufert, Guggemos, & Sailer, 2021).

6.3. Implications

Kirschner and van Merriënboer (2013) raised concerns about whether students know best about what is conducive to their learning. Indeed, our findings show that teachers may be misguided if they give their students a choice whether to use (further) K-12 MOOCs in their instruction, when seeking IL gain. Student assessment may be purely based on whether they regard the MOOC as fun and entertaining (hedonic motivation); the acceptance of our MOOC can be almost perfectly predicted with hedonic motivation. However, hedonic motivation is a negative predictor for IL gain. If the student perspective is to be considered, it may be more revealing to ask them about their performance expectancy and effort expectancy as these are positive predictors for IL gain. In light of this, our research may help teachers and educational decision makers to ask students appropriate questions when deciding about integrating a MOOC at the K-12 level.

MOOCs are an important phenomenon in higher education; however, our study shows that upper secondary level students may have detrimental perceptions about them. It may be important to help students form realistic perceptions about MOOCs. According to Kirschner and Bruyckere (2017), it may be a myth that students can use technology, in our case a MOOC, without thorough instruction. A key variable in the instructional design is the level of guidance and support (Kirschner, Sweller, & Clark, 2006). When MOOCs are introduced at the K-12 level, teachers may provide their students with much guidance and support to help them acquire the skills conducive to successfully performing in the course. These skills could be (Castaño-Muñoz & Rodrigues, 2021): organization skills, self-regulation skills, and digital skills, e.g., digital collaboration and communication. The guidance and support may fade as students become more familiar with the use of MOOCs. Teachers may also point out that using a K-12 MOOC is not all about fun and entertainment (hedonic motivation). Rather, the successful completion of a MOOC could be rewarding (performance expectancy). Overall, a deliberate introduction of MOOCs at the K-12 level could foster the skills and attitudes that are necessary for successfully completing MOOCs. This may also yield benefits beyond higher education. MOOCs could play a vital role in workplace learning (Egloffstein & Ifenthaler, 2017). However, the current situation, where mostly highly qualified professionals use MOOCs, might be suboptimal

(Castaño-Muñoz & Rodrigues, 2021). A good example of the integration of MOOCs into a lifelong learning process is the initiative Digital Israel (https://campus.gov.il/en/about/): teachers in secondary education are required to use at least one MOOC during a semester in order to prepare students for this lifelong process (Seufert, Guggemos, & Moser, 2019).

As noted in the limitation section, our outcome variables, technology acceptance, and IL learning are narrow in scope. Positive side effects concerning the increase of digital skills as conceptualized by the DigComp 2.2 framework (Vuorikari, Kluzer, & Punie, 2022) are likely. As can be seen from the supplementary material S2, students communicate and collaborate to carry out the tasks, create digital content, e.g., videos, manage safety, e.g., use strong passwords, and solve digital problems, e.g., handling troubles when using the MOOC platform. Teachers could use K-12 MOOCs as a valuable learning resource to address important digital skills beyond the content matter (in the case of the i-MOOC, information literacy). While doing so, they may also improve their own technology-related knowledge and skills, e.g., in the realm of learning analytics (Ifenthaler, Gibson, Prasse, Shimada, & Yamada, 2021).

Credit author statement

Josef Guggemos: Conceptualization, Methodology, Validation, Formal analysis (statistical analysis), Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition, Project administration, Funding acquisition, **Luca Moser**: Investigation, Validation, Formal analysis (literature review), Data curation, Writing – original draft, Writing – review & editing, Project administration, **Sabine Seufert**: Conceptualization, Writing – review & editing, Resources, Supervision, Funding acquisition.

Declaration of competing interest

None.

Acknowledgments

The underlying project was funded by the Swiss National Science Foundation (no. 178509). The funder has no influence on the study design, data collection, analysis and interpretation of data, writing of the report, and the decision to submit the article for publication. We thank Katharina Scheiter for her thoughtful comments and suggestions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.compedu.2022.104552.

References

Aichele, C., Hartig, J., & Michaelis, C. (2021). Assessing learning progress: Validating a test score interpretation in the domain of sustainability management. *Studies in Higher Education, 46*(10), 2047–2062. https://doi.org/10.1080/03075079.2021.1953329

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2018). Technology acceptance model in M-learning context: A systematic review. *Computers & Education, 125*, 389–412. https://doi.org/10.1016/j.compedu.2018.06.008

BBT. (2006). *Berufliche Grundbildung: Rahmenlehrplan für den allgemeinbildenden Unterricht [Basic vocational education: Framework curriculum for general education classes]*. Bundesamt für Berufsbildung und Technologie (BBT). https://www.svabu.ch/public/042/01/RLP_ABU.pdf.

Bernath, J., Suter, L., Waller, G., Külling, C., Willemse, I., & Süss, D. (2020). James: Jugend, Aktivitäten, Medien – Erhebung Schweiz [Youth, activities, media: Survey Switzerland]. *ZHAW Zürcher Hochschule für Angewandte Wissenschaften*. https://doi.org/10.21256/zhaw-21175

Blazquez-Merino, M., Macho-Aroca, A., Baizán-Alvarez, P., Garcia-Loro, F., Cristobal, E. S., Diez, G., et al. (2018). Structured MOOC designed to optimize electricity learning at secondary school. In *2018 IEEE global engineering education conference (EDUCON)*. https://doi.org/10.1109/EDUCON.2018.8363232

Bralić, A., & Divjak, B. (2018). Integrating MOOCs in traditionally taught courses: Achieving learning outcomes with blended learning. *International Journal of Educational Technology in Higher Education, 15*(1), 1–16. https://doi.org/10.1186/s41239-017-0085-7

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Expanded edition). National Academy Press. https://doi.org/10.17226/9853

van Breukelen, G. J. P. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research, 48*(6), 895–922. https://doi.org/10.1080/00273171.2013.831743

Briggs, S., & Crompton, H. (2016). Taking advantage of MOOCs in K-12 education: A blended approach. In *Mobile and blended learning innovations for improved learning outcomes* (pp. 297–309). IGI Global. https://doi.org/10.4018/978-1-5225-0359-0.ch015.

Cachero, C., Barra, P., Melia, S., & Lopez, O. (2020). Impact of programming exposure on the development of computational thinking capabilities: An empirical study. *IEEE Access, 8*, 72316–72325. https://doi.org/10.1109/ACCESS.2020.2987254

Canessa, E., & Pisani, A. (2013). High school open on-line courses (HOOC): A case study from Italy. *European Journal of Open, Distance and E-Learning, 16*(1). https://old.eurodl.org/materials/contrib/2013/Canessa_Pisani.pdf.

Castaño-Muñoz, J., & Rodrigues, M. (2021). Open to MOOCs? Evidence of their impact on labour market outcomes. *Computers & Education, 173*, 104289. https://doi.org/10.1016/j.compedu.2021.104289

Chen, X., Zou, D., Xie, H., & Zou, Di (2020). Fifty years of British journal of educational technology: A topic modeling based bibliometric perspective. *British Journal of Educational Technology, 51*(3), 692–708. https://doi.org/10.1111/bjet.12907

Cohen, L., & Magen-Nagar, N. (2016). Self-regulated learning and a sense of achievement in MOOCs among high school science and technology students. *American Journal of Distance Education, 30*(2), 68–79. https://doi.org/10.1080/08923647.2016.1155905

de Waard, I., & Demeulenaere, K. (2017). *The MOOC-CLIL project: Using MOOCs to increase language, and social and online learning skills for 5th grade K-12 students* (pp. 29–42). Dublin, Ireland: Beyond the Language Classroom: Researching MOOCs and Other Innovations.

Deutscher, V., & Winther, E. (2018). Instructional sensitivity in vocational education. *Learning and Instruction, 53*, 21–33. https://doi.org/10.1016/j.learninstruc.2017.07.004

DiStefano, C., Zhu, M., & Mîndrilã, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research and Evaluation, 14*(20), 1–11. https://doi.org/10.7275/DA8T-4G52

Dziabenko, O., & Tsourlidaki, E. (2018). MOOC in a school environment: ODL project. In M. E. Auer, & D. G. Zutin (Eds.), *Online engineering & internet of things* (pp. 833–839). Springer International Publishing.

EDK. (1994). *Rahmenlehrplan für die Maturitätsschulen [Framework curriculum for high schools]*. https://edudoc.ch/record/17476/files/D30a.pdf.

Egloffstein, M., & Ifenthaler, D. (2017). Employee perspectives on MOOCs for workplace learning. *TechTrends, 61*(1), 65–70. https://doi.org/10.1007/s11528-016-0127-3

Egloffstein, M., Koegler, K., & Ifenthaler, D. (2019). Instructional quality of business MOOCs: Indicators and initial findings. *Online Learning Journal, 23*(4), 85–105. https://doi.org/10.24059/olj.v23i4.2091

Filvà, D. A., Guerrero, M. J. C., & Forment, M. A. (2014). The effects of massiveness on the participation in social technologies: A MOOC in secondary education. In *Proceedings of the second international conference on technological ecosystems for enhancing multiculturality* (pp. 397–402). Association for Computing Machinery (ACM). https://doi.org/10.1145/2669711.2669930.

Finney, S. J., Sundre, D. L., Swain, M. S., & Williams, L. M. (2016). The validity of value-added estimates from low-stakes testing contexts: The impact of change in test-taking motivation and test consequences. *Educational Assessment, 21*(1), 60–87. https://doi.org/10.1080/10627197.2015.1127753

Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). *International computer and information literacy study 2018: Assessment framework*. Amsterdam: IEA. https://www.iea.nl/studies/iea/icils/2018.

Grella, T. C., Staubitz, T., & Meinel, C. (2016). Can MOOCs support secondary education in computer science?. In M. E. Auer, D. Guralnick, & J. Uhomibhi (Eds.), *Interactive collaborative learning: Proceedings of the 19th ICL conference* (Vol. 544, pp. 478–493) IEEE. https://doi.org/10.1007/978-3-319-50337-0_45.

Grover, S., Pea, R. D., & Cooper, S. (2016). Factors influencing computer science learning in middle school. In *Proceedings of the 47th ACM technical symposium on computing science education* (pp. 552–557). https://doi.org/10.1145/2839509.2844564

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hermans, F., & Aivaloglou, E. (2017). Teaching software engineering principles to K-12 students: A MOOC on scratch. In *In 2017 IEEE/ACM 39th international conference on software engineering: Software engineering education and training track (ICSE-SEET)*. https://doi.org/10.1109/ICSE-SEET.2017.13

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education, 98*, 157–168. https://doi.org/10.1016/j.compedu.2016.03.016

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A., & Yamada, M. (2021). Putting learning back into learning analytics: Actions for policy makers, researchers, and practitioners. *Educational Technology Research & Development, 69*(4), 2131–2150. https://doi.org/10.1007/s11423-020-09909-8

Israel, M. J. (2015). Effectiveness of integrating MOOCs in traditional classrooms for undergraduate students. *International Review of Research in Open and Distance Learning, 16*(5), 102–118. https://eric.ed.gov/?id=EJ1077803.

Janisch, S., Ebner, M., & Slany, W. (2017). Informatische Bildung mithilfe eines MOOC. *Erziehung & Unterricht, 167*, 7–8. https://graz.pure.elsevier.com/en/publications/informatische-bildung-mithilfe-eines-mooc.

Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in massive open online courses. *Computers & Education, 146*, 103771. https://doi.org/10.1016/j.compedu.2019.103771

Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods, 51*(5), 2337–2355. https://doi.org/10.3758/s13428-018-1187-4

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., et al. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research, 88*(1), 43–86. https://doi.org/10.3102/0034654317740335

Kasch, J., Van Rosmalen, P., & Kalz, M. (2021). Educational scalability in MOOCs: Analysing instructional designs to find best practices. *Computers & Education, 161*, 104054. https://doi.org/10.1016/j.compedu.2020.104054

de Kereki, I. F., & Manataki, A. (2016). "Code yourself" and "A programar": A bilingual MOOC for teaching computer science to teenagers. In *2016 IEEE frontiers in education conference (FIE)*. https://doi.org/10.1109/FIE.2016.7757569

de Kereki, I. F., & Paulos, V. (2014). SM4T: Scratch MOOC for teens: A pioneer pilot experience in Uruguay. In *IEEE Frontiers in Education Conference (FIE) Proceedings (pp. 1–4)*. IEEE. https://doi.org/10.1109/FIE.2014.7044264.

Khalil, M., & Ebner, M. (2015). A STEM MOOC for school children — what does learning analytics tell us?. In *2015 international conference on interactive collaborative learning (ICL)* (pp. 1217–1221). IEEE. https://doi.org/10.1109/ICL.2015.7318212.

Khechine, H., Raymond, B., & Augier, M. (2020). The adoption of a social learning system: Intrinsic value in the UTAUT model. *British Journal of Educational Technology, 51*(6), 2306–2325. https://doi.org/10.1111/bjet.12905

Kim, D., Jung, E., Yoon, M., Chang, Y., Park, S., Kim, D., & Demir, F. (2021). Exploring the structural relationships between course design factors, learner commitment, self-directed learning, and intentions for further learning in a self-paced MOOC. *Computers & Education, 166*, 104171. https://doi.org/10.1016/j.compedu.2021.104171

Kirschner, P. A., & Bruyckere, P. de (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education, 67*, 135–142. https://doi.org/10.1016/j.tate.2017.06.001

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kirschner, P. A., & van Merriënboer, J. J. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist, 48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Kirschner, P. A., Verschaffel, L., Star, J., & van Dooren, W. (2017). There is more variation within than across domains: An interview with Paul A. Kirschner about applying cognitive psychology-based instructional design principles in mathematics teaching and learning. *ZDM, 49*(4), 637–643. https://doi.org/10.1007/s11858-017-0875-3

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education, 104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kohl, M. (2020). *Package 'MKinfer'*. https://cran.r-project.org/web/packages/MKinfer/MKinfer.pdf.

Köhler, C., Pohl, S., & Carstensen, C. H. (2017). Dealing with item nonresponse in large-scale cognitive assessments: The impact of missing data methods on estimated explanatory relationships. *Journal of Educational Measurement, 54*(4), 397–419. https://doi.org/10.1111/jedm.12154

Koutsakas, P., Chorozidis, G., Karamatsouki, A., & Karagiannidis, C. (2020). Research trends in K–12 MOOCs: A review of the published literature. *International Review of Research in Open and Distance Learning, 21*(3), 285–303. https://doi.org/10.19173/irrodl.v21i3.4650

Koutsakas, P., Karagiannidis, C., Politis, P., & Karasavvidis, I. (2020). A computer programming hybrid MOOC for Greek secondary education. *Smart Learning Environments, 7*(1), 1–22. https://doi.org/10.1186/s40561-020-0114-1

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review, 25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Kurhila, J., & Vihavainen, A. (2015). A purposeful MOOC to alleviate insufficient CS education in Finnish schools. *ACM Transactions on Computing Education, 15*(2). https://doi.org/10.1145/2716314

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology, 32*(1). https://doi.org/10.5334/irsp.289

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, L., Johnson, J., Aarhus, W., & Shah, D. (2022). Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit. *Computers & Education, 176*, 104354. https://doi.org/10.1016/j.compedu.2021.104354

Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Journal of Educational Technology & Society, 21*(2), 220–232. https://eric.ed.gov/?id=EJ1175301.

Magen-Nagar, N., & Cohen, L. (2017). Learning strategies as a mediator for motivation and a sense of achievement among students who study in MOOCs. *Education and Information Technologies, 22*(3), 1271–1290. https://doi.org/10.1007/s10639-016-9492-y

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Comminfolit, 10*(2), 199–213. https://doi.org/10.15760/comminfolit.2016.10.2.24

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education, 80*, 77–83. https://doi.org/10.1016/j.compedu.2014.08.005

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*(1), 87–91. https://doi.org/10.1119/1.2372468

McGill, T. J., Klobas, J. E., & Renzi, S. (2014). Critical success factors for the continuation of e-learning initiatives. *The Internet and Higher Education, 22*, 24–36. https://doi.org/10.1016/j.iheduc.2014.04.001

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24*(1), 20–35. https://doi.org/10.1037/met0000182

McNeish, D., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*(2), 295–314. https://doi.org/10.1007/s10648-014-9287-x

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140. https://doi.org/10.1037/met0000078

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods, 52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Meet, R. K., & Kala, D. (2021). Trends and future prospects in MOOC researches: A systematic literature review 2013-2020. *Contemporary Educational Technology, 13*(3), Article ep312. https://doi.org/10.30935/cedtech/10986

Moser, L., Guggemos, J., & Seufert, S. (2021). Improving a MOOC to foster information literacy by means of a conjecture map. *International Journal of Learning Technology, 16*(1), 65–86. https://doi.org/10.1504/IJLT.2021.115470

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 54–60. https://doi.org/10.1080/10705511.2014.950896

de Moura, V. F., de Souza, C. A., Viana, A. B. N., Feitosa de Moura, V., Alexandre de Souza, C., & Noronha Viana, A. B. (2021). The use of Massive Open Online Courses (MOOCs) in blended learning courses and the functional value perceived by students. *Computers & Education, 161*, 104077. https://doi.org/10.1016/j.compedu.2020.104077

Muñoz-Merino, P. J., Valiente, J. A. R., & Kloos, C. D. (2013). Inferring higher level learning information from low level data for the Khan Academy platform. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 112–116)*. ACM Press. https://doi.org/10.1145/2460296.2460318. ACM.

Najafi, H., Evans, R., & Federico, C. (2014). MOOC integration into secondary school courses. *International Review of Research in Open and Distance Learning, 15*(5). https://doi.org/10.19173/irrodl.v15i5.1861

Naumann, A., Rieser, S., Musow, S., Hochweber, J., & Hartig, J. (2019). Sensitivity of test items to teaching quality. *Learning and Instruction, 60*, 41–53. https://doi.org/10.1016/j.learninstruc.2018.11.002

Nigh, J., Pytash, K. E., Ferdig, R. E., & Merchant, W. (2015). Investigating the potential of MOOCs in K-12 teaching and learning environments. *Journal of Online Learning Research, 1*(1), 85–106. https://www.learntechlib.org/primary/p/149853/.

Novak, E. (2021). *Mathematical modeling for theory-oriented research in educational technology. Educational Technology Research and Development*. Advance online publication. https://doi.org/10.1007/s11423-021-10069-6

Oberländer, M., Beinicke, A., & Bipp, T. (2020). Digital competencies: A review of the literature and applications in the workplace. *Computers & Education, 146*, 103752. https://doi.org/10.1016/j.compedu.2019.103752

Panyajamorn, T., Kohda, Y., Chongphaisal, P., & Supnithi, T. (2016). The effectiveness and suitability of MOOCs hybrid learning: A case study of public schools in Thai rural area. In *In 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS) (pp. 1–6)*. IEEE. https://doi.org/10.1109/KICSS.2016.7951449.

Paton, R. M., Scanlan, J. D., & Fluck, A. E. (2018). A performance profile of learner completion and retention in Australian VET MOOCs. *Journal of Vocational Education and Training, 70*(4), 581–599. https://doi.org/10.1080/13636820.2018.1463278

Perach, S., & Alexandron, G. (2021). *A Mooc-based computer science program for middle school: Results, challenges, and the Covid-19 effect*. The Seventh European MOOCs Stakeholder Summit (EMOOCs 2021). https://easychair.org/publications/preprint/1gC9.

Politis, P., Koutsakas, P., & Karagiannidis, C. (2017). A MOOC for secondary education in Greece. In *In 1st International Conference on Smart Learning for Community Development (pp. 134-147)*. https://dspace.qou.edu/contents/smart/resources/documents/smart_learning.pdf.

Popham, W. J. (2001). Teaching to the test? *Educational Leadership, 58*(6), 16–20. https://eric.ed.gov/?id=EJ626269.

Raffaghelli, J. E., Rodríguez, M. E., Guerrero-Roldán, A.-E., & Bañeres, D. (2022). Applying the UTAUT model to explain the students' acceptance of an early warning system in Higher Education. *Computers & Education, 182*, 104468. https://doi.org/10.1016/j.compedu.2022.104468

Ranellucci, J., Rosenberg, J. M., & Poitras, E. G. (2020). Exploring pre-service teachers' use of technology: The technology acceptance model and expectancy-value theory. *Journal of Computer Assisted Learning, 36*(6), 810–824. https://doi.org/10.1111/jcal.12459

Reeves, T. C., & Lin, L. (2020). The research we have is not the research we need. *Educational Technology Research & Development, 68*(4), 1991–2001. https://doi.org/10.1007/s11423-020-09811-3

Reich, J. (2015). Education research. Rebooting MOOC research. *Science, 347*(6217), 34–35. https://doi.org/10.1126/science.1261627

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science, 363*(6423), 130–131. https://doi.org/10.1126/science.aav7958

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction, 21*(4), 550–559. https://doi.org/10.1016/j.learninstruc.2010.11.001

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–35. https://doi.org/10.18637/jss.v048.i02

Sands, P., & Yadav, A. (2020). Self-regulation for high school learners in a MOOC computer science course. In *Proceedings of the 51st ACM technical symposium on computer science education* (pp. 845–851). Association for Computing Machinery. https://doi.org/10.1145/3328778.3366818.

Savalei, V. (2019). A comparison of several approaches for controlling measurement error in small samples. *Psychological Methods, 24*(3), 352–370. https://doi.org/10.1037/met0000181

SBFI. (2013). *Rahmenlehrplan für die Berufsmaturität [Framework curriculum for vocational high schools]*. Schweiz. *Staatssekretariat für Bildung, Forschung und Innovation (SBFI)*. https://edudoc.ch/record/109355?ln=de.

Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung [Technology-enhanced learning and teaching: An overview]. *Zeitschrift fur Erziehungswissenschaft, 1–22*. https://doi.org/10.1007/s11618-021-01047-y

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education, 128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Seufert, S., Guggemos, J., & Moser, L. (2019). Digitale Transformation in Hochschulen: Auf dem Weg zu offenen Ökosystemen [Digital transformation in higher education: Towards open ecosystems]. *Zeitschrift für Hochschulentwicklung, 14*(2), 85–107. https://doi.org/10.3217/zfhe-14-02/05

Seufert, S., Guggemos, J., Moser, L., & Sonderegger, S. (2019). Developing a MOOC to foster information literacy by means of a conjecture map. *Communications in Computer and Information Science, 1011*, 202–213. https://doi.org/10.1007/978-3-030-20798-4_18

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior, 115*. https://doi.org/10.1016/j.chb.2020.106552

Seufert, S., Stanoevska-Slabeva, K., & Guggemos, J. (2020). Assessing subjective and objective information literacy at upper secondary schools – an empirical study in four German-speaking countries. *International Journal of Learning Technology, 15*, 82–103. https://doi.org/10.1504/IJLT.2020.107666

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*(4), 563–575. https://doi.org/10.1007/BF02296196

Staubitz, T., Teusner, R., & Meinel, C. (2019). MOOCs in secondary education - experiments and observations from German classrooms. In *2019 IEEE global engineering education conference (EDUCON)*. https://doi.org/10.1109/EDUCON.2019.8725138

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research & Development, 68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tomkins, S., & Getoor, L. (2019). Understanding hybrid-MOOC effectiveness with a collective socio-behavioral model. *Journal of Educational Data Mining, 11*(3), 42–77. https://doi.org/10.5281/zenodo.3594773

Tomkins, S., Ramesh, A., & Getoor, L. (2016). Predicting post-test performance from online student behavior: A high school MOOC case study. In *Proceedings of the 9th international conference on educational data mining* (pp. 239–246). https://www.educationaldatamining.org/EDM2016/proceedings.html.

Valtonen, T., López-Pernas, S., Saqr, M., Vartiainen, H., Sointu, E. T., & Tedre, M. (2022). The nature and building blocks of educational technology research. *Computers in Human Behavior, 128*, 107123. https://doi.org/10.1016/j.chb.2021.107123

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., Xu, X., Venkatesh, T., & Xu. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Wang, X., Hall, A. H., & Wang, Q. (2019). Investigating the implementation of accredited massive online open courses (MOOCs) in higher education: The boon and the bane. *Australasian Journal of Educational Technology, 35*(3). https://doi.org/10.14742/ajet.3896

Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science, 4*(1), 1–17. https://doi.org/10.1177/2515245920918253

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes*. European Commission. https://doi.org/10.2760/490274.

Wang, K., & Zhu, C. (2019). MOOC-based flipped learning in higher education: Students' participation, experience and learning performance. *International Journal of Educational Technology in Higher Education, 16*(1), 1–18. https://doi.org/10.1186/s41239-019-0163-0

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495. https://doi.org/10.5465/amr.1989.4308371

Wigfield, A., Muenks, K., & Eccles, J. S. (2021). Achievement motivation: What we know and where we are going. *Annual Review of Developmental Psychology, 3*(1), 87–111. https://doi.org/10.1146/annurev-devpsych-050720-103500

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934. https://doi.org/10.1177/0013164413495237

Yang, S., Lee, J. W., Kim, H.-J., Kang, M., Chong, E., & Kim, E. (2021). Can an online educational game contribute to developing information literate citizens? *Computers & Education, 161*, 104057. https://doi.org/10.1016/j.compedu.2020.104057

Yin, Y., Adams, C., Goble, E., & Vargas-Madriz, L. F. (2015). A classroom at home: Children and the lived world of MOOCs. *Educational Media International, 52*(2), 88–99. https://doi.org/10.1080/09523987.2015.1053287

Yousef, A. M. F., & Sumner, T. (2021). Reflections on the last decade of MOOC research. *Computer Applications in Engineering Education, 29*(4), 648–665. https://doi.org/10.1002/cae.22334

Zhu, M., Sari, A. R., & Lee, M. M. (2020). A comprehensive systematic review of MOOC research: Research techniques, topics, and trends from 2009 to 2019. *Educational Technology Research & Development, 68*(4), 1685–1710. https://doi.org/10.1007/s11423-020-09798-x

Studie 5:

Guggemos, J., Seufert, S. & Sonderegger, S. (2020). Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology*, 51(5), 1864–1883.

British Journal of Educational Technology Vol 0 No 0 2020 1–2
doi:10.1111/bjet.13006

Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT

Josef Guggemos , Sabine Seufert and Stefan Sonderegger

Josef Guggemos is a postdoctoral researcher at the Institute of Business Education and Educational Management, University of St. Gallen, Switzerland. He obtained a PhD at the Munich School of Management, LMU Munich, Germany. His research focus includes computational thinking, teachers' competencies in the context of digital transformation and social robots in education. Sabine Seufert is a full professor and director of the Institute of Business Education and Educational Management, University of St. Gallen, Switzerland. Her research is situated within the digital transformation of education. She is a member of the editorial board of the International Journal of Learning Technology and the International Journal of Advanced Corporate Learning. Stefan Sonderegger is a research and teaching assistant at the Institute of Business Education and Educational Management, University of St. Gallen, Switzerland. His research concerns social robots in education and learning analytics. Address for correspondence: Dr Josef Guggemos, Institute of Business Education and Educational Management, University of St.Gallen, St. Jakob-Str. 21, St. Gallen 9000, Switzerland. Email: josef.guggemos@unisg.ch

Abstract

This study investigates the acceptance of social robots by higher education students in the social sciences. Pepper, a humanoid social robot from SoftBank Robotics, provided a sample of its capabilities during a first semester, large-scale, university course, "Introduction to academic writing."From this course, 462 freshmen participated in our survey. The unified theory of acceptance and use of technology (UTAUT) acts as the conceptual framework, and partial least squares structural equation modelling (PLS-SEM) as the method for data analysis. The four perceived characteristics—trustworthiness, adaptiveness, social presence and appearance—all predict the intention to use the robot for learning purposes; anxiety regarding making mistakes in handling the robot and about privacy issues are not significant predictors. An importance-performance map analysis indicated adaptiveness as the robot's most important characteristic for predicting student behavioural intention. Overall, however, the study shows that students do not have the intention to rely on social robots for learning purposes at the current level of state-of-the-art technology: behavioural intention reaches only 36.6% of the theoretical maximum.

Introduction

Social robots have the potential to become an integral part of the educational infrastructure (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018; Mubin, Stevens, Shahid, Mahmud, & Dong, 2013): as tutors or teachers, they might take over selected duties that human teachers carry out. For instance, they can provide hints upon request to students. By assisting the human teacher, they free up resources, ie, comparative advantages can be realised. As peers, they are companions of students and collaboratively learn with them. In this case, they could help to keep student motivation at a high level by encouraging them. Social robots can also play the role

Practitioner Notes

What is already known about this topic?

- Most studies on social robots have focused on children; in higher education, social robots have most often been used in courses on computer science and robotics.
- The majority of research concerning social robots in higher education has been carried out in the form of literature reviews; empirical evidence is regularly missing.
- The UTAUT is suitable for evaluating the acceptance of social robots.
- Nao, a humanoid social robot from SoftBank Robotics, is a quasi-standard type used in education; in commercial settings its successor, Pepper, is more widely utilised.

What this paper adds

- An investigation into the acceptance of social robots for educational purposes in higher education, specifically in a large-scale academic writing course.
- Evidence concerning the association between privacy concerns and the acceptance of social robots for learning purposes.
- Use of PLS-SEM, including importance-performance map and multigroup analyses, to provide an in-depth insight into student acceptance of social robots.
- The most important characteristic of the robot in predicting its acceptance for learning purposes is adaptiveness.

Implications for practice and/or policy

- To achieve a high degree of adaptiveness, much effort may be required in developing the underlying artificial intelligence. Research on artificial intelligence (the basis for adaptiveness) and social robotics may inform each other.
- The voice (acoustical appearance) of the robot, rather than its visual appearance, is more important in forming its overall appearance.
- Currently, student intention to rely on social robots for learning purposes is at a rather low level; our study contributes to a better understanding of how this intention could be increased.

of novice, ie, students are asked to explain something to the robot. The rationale behind this is a learning-by-teaching method.

A social robot is characterised by a physical presence, which distinguishes it from virtual agents that might appear on a screen. Since the physical presence incurs additional expenses, the use of social robots instead of virtual agents has to be justified (Belpaeme *et al.*, 2018). The expenses involved are dependent on the type of robot. Lehmann and Rossi (2019) provide an overview of available models. Usually, social robots are humanoid; however, some of them are zoomorphic. Nao (see Figure 1), developed by SoftBank Robotics with a height of 54 cm, may be a quasi-standard form in education (Belpaeme *et al.*, 2018). In commercial settings, Pepper (see Figure 1), also a humanoid robot from SoftBank Robotics, is more widely used (Gardecki & Podpora, 2017). Examples of places where Pepper provides information to visitors are banks and shopping malls. Its height of 120 cm and overall appearance has been designed based on studies regarding the acceptance of robots as assistants (Pandey & Gelin, 2018). Pepper can communicate verbally and nonverbally, eg, using speech, gesture and gaze (Huang & Mutlu, 2014). Moreover, a tablet placed on Pepper's chest can be used for receiving user input and for presenting output. Besides commercial settings, Pepper has also been utilised in education, eg, for language learning in

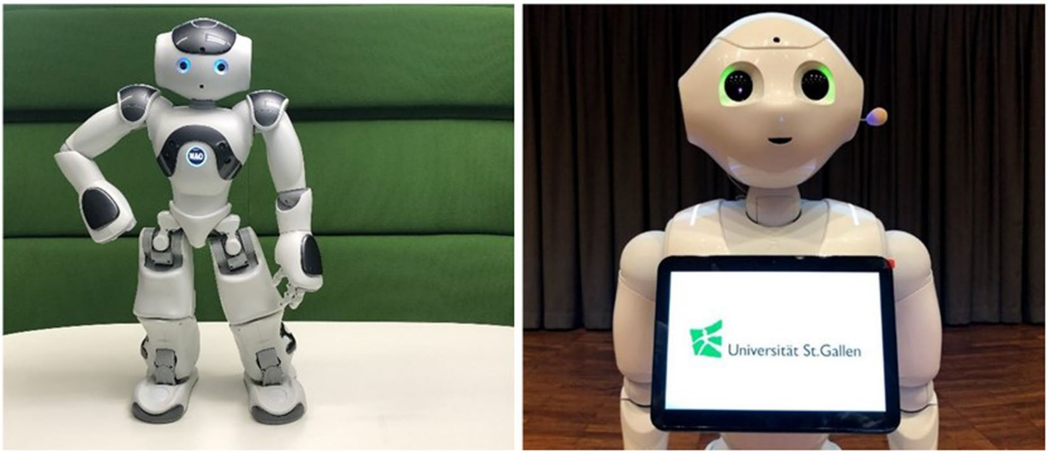


Figure 1: On the left, social robot "Nao"; on the right, "Pepper"

children (Tanaka *et al.*, 2015). Due to the described characteristics, Pepper may also be a valuable resource in higher education. Lehmann and Rossi (2019) developed the concept of a 180-minute lecture in which Pepper is an integral part. The robot introduces the lesson, offers short games to the students during the break, presents the results of a quiz carried out with Google forms and summarises the lesson. The concept aims at increasing the student engagement using Pepper to provide feedback. However, empirical evidence on the outcome of this project is currently not available. In general, evidence concerning social robots in higher education is scarce (Spolaôr & Benitti, 2017; Zhong & Xia, 2018). Studies have been mainly carried out in the fields of computer science (Abildgaard & Scharfe, 2012; Byrne, Rossi, & Doolan, 2017) and robotics (Bacivarov & Ilian, 2012; Bolea, Grau, & Sanfeliu, 2016; Gao, Barendregt, Obaid, & Castellano, 2018).

Social robots could be a valuable resource, especially in large-scale, university courses (Byrne, Rossi, & Doolan, 2017; Cooney & Leister, 2019; Handke, 2018). In such an environment it is often hard to adequately support students and answer individual questions. Employing human assistants for this purpose may not be a solution for various reasons, mainly budget restrictions. Research is necessary to gain a better understanding of social robots as teaching assistants in higher education courses (Handke, 2018). To this end, for our study we utilise an academic writing course with more than 1500 students. This setting is characterised by significant heterogeneity, in terms of students' prior knowledge (Seufert & Spiroudis, 2017). After an introduction lecture, students are assigned to small groups. For the success of this group work, a common understanding is important, eg, in terms of the evaluation criteria and the citation standards. Social robots could be useful in this process by offering further assistance. A prerequisite for a successful integration, however, may be technology acceptance (Venkatesh, Morris, Davis, & Davis, 2003). If students were unwilling to utilise social robots for learning purposes they would not be useful, regardless of their actual value.

The unified theory of acceptance and use of technology (UTAUT) is regularly used to evaluate the acceptance of technology in educational contexts (eg, Garone *et al.*, 2019; Nistor, Stanciu, Lerche, & Kiel, 2019). Fridin and Belokopytov (2014) demonstrated, by means of a literature review, the suitability of the UTAUT for evaluating the acceptance of social robots. In higher education, Conti, Cattani, Di Nuovo and Di Nuovo (2019) relied on the UTAUT to evaluate the acceptance of a Nao-type model. They assessed the intention of prospective psychologists to use the robot as a tool in their future clinical practice.

Theoretical background of evaluation

The UTAUT (Venkatesh *et al.*, 2003) claims that behavioural intention (BI) can be directly predicted by performance expectancy (PE), effort expectancy (EE) and social influence (SI). Adapted to our context, we define BI as the intention to use a social robot for learning purposes, PE as the degree to which students believe that using social robots will be helpful in their learning process and EE as the degree of ease associated with the use of social robots in terms of learning. SI is defined as the degree to which students assume that others believe they should use social robots for learning purposes.

The UTAUT proposes that variables beyond PE, EE and SI, eg, attitudes and anxiety, have no direct influence on BI. Hence, we are interested in how perceived characteristics of the robot can predict PE and EE and, hence, indirectly BI. The following section presents characteristics that have been identified as important for the acceptance of social robots in previous studies.

Trust plays a decisive role in human–robot interaction (Hancock *et al.*, 2011). A robot may be regarded as trustworthy if it shows an acceptable level of competence and integrity (Devitt, 2018). From a technology acceptance perspective, there is evidence for a positive influence of trust on PE and EE (Lee & Song, 2013).

Adaptiveness is regarded as a key facet of quality teaching (Brühwiler & Blatchford, 2011). Concerning social robots in the context of learning, perceived adaptiveness may be conceptualised as the degree to which students believe the robot takes into account their personal (learning) needs. There is evidence for adaptiveness being a positive influence on PE (Fridin & Belokopytov, 2014; Graaf & Allouch, 2013; Heerink, Kröse, Evers, & Wielinga, 2010). We also assume a positive influence of perceived adaptiveness on EE: a robot that adapts to a higher degree to individual needs may also be easier to use.

As the name implies, the main advantage of social robots in comparison to virtual agents may be social presence, which can be defined as “the experience of sensing a social entity when interacting with the system” (Heerink *et al.*, 2010, p. 364). There is evidence for social presence being a positive influence on attitudes regarding the use of robots (Graaf & Allouch, 2013). Shin and Choo (2011) showed that social presence is a key factor in the acceptance of social robots. However, Kennedy, Baxter and Belpaeme (2015), relying on a sample of children aged seven to eight, conclude that social presence could also be detrimental because it can detract students from the content. In either case, social presence may be a predictor for PE and EE.

The appearance of a social robot is determined by its size, shape, look and voice: characteristics that play a crucial role in the design process (Pandey & Gelin, 2018). Research in the realm of human–robot interaction indicates the robot’s appearance to be an important influence on its perception by users (Tapus, Mataric, & Scassellati, 2007; Walters, Syrdal, Dautenhahn, te Boekhorst, & Koay, 2008). Against this backdrop, we expect the appearance of the robot to predict PE and EE.

Besides the characteristics of social robots, anxiety about them should be taken into account (Graaf & Allouch, 2013; Heerink *et al.*, 2010). Following the UTAUT, the influence of anxiety on BI is fully mediated by EE. We, therefore, expect anxiety to negatively predict EE. In terms of social robots, anxiety might comprise two facets. Moreover, students could be afraid of making mistakes or breaking something when handling the robot (Heerink *et al.*, 2010); on the other, they could

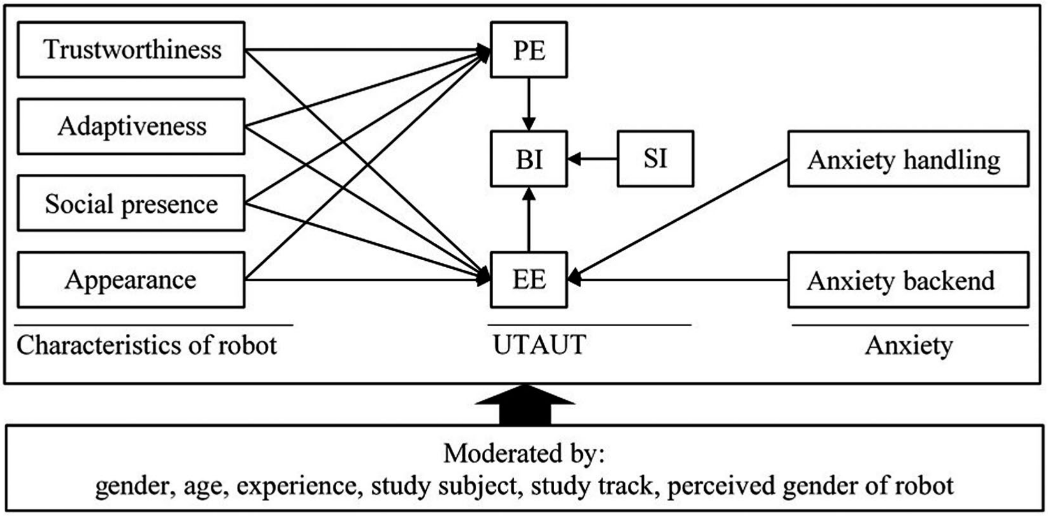


Figure 2: Conceptual framework for predicting Behavioural Intention (BI)

be worried about what happens behind the scenes. When interacting with social robots, students may be concerned about informational privacy (Lutz, Schöttler, & Hoffmann, 2019), eg, what happens with private information collected by the robot. With this in mind, such concerns, ie, anxiety about what happens in the backend, should be considered when investigating the acceptance of social robots.

The UTAUT implies that relationships may be moderated by gender, age and experience (Venkatesh *et al.*, 2003). Besides these generic variables, further moderators may be considered due to the context of our study and evidence from robot research. The way the robot is viewed may vary between students of different study subjects and tracks. Another potential moderator is the perceived gender—a characteristic that influences how a robot is regarded (Appel, Izydorczyk, Weber, Mara, & Lischetzke, 2020). Figure 2 summarises our conceptual framework for predicting the intention to use social robots in higher education.

In line with Fridin and Belokopytov (2014), as well as Conti *et al.* (2019), we use BI as our outcome variable. The research questions that we aim to answer are:

RQ1: What predicts the intention of students in higher education to use social robots for learning purposes (BI)?

RQ2: Which characteristics of the robot predict BI?

RQ3: How does anxiety predict BI?

Study design

The study was carried out during the first lecture of the large-scale course, "Introduction to academic writing," which is mandatory for all 1552 freshmen at the University of St.Gallen. The course has an English track (470 students) and a German one (1082 students). Therefore, the lecture was conducted in both English and German. In the lecture, we aimed to offer the students a representative sample of tasks that a social robot might perform within the context of learning at the current level of state-of-the-art technology. To this end, we drew on the work of Cooney and

Leister (2019), who provided a list of tasks that social robots could carry out in higher education. Based on this, we developed samples that may fairly represent the capabilities of the social robot. Besides generic activities like greeting, we relied on our own experience regarding typical issues and questions in academic writing, eg, how plagiarism is detected. The robot should demonstrate its ability to answer such typical questions. To ensure a state-of-the-art technical implementation, we collaborated with raumCode from Zurich, a company specialising in humanoid robots and artificial intelligence. Our part in this collaboration was to set the content and pedagogical-related goals. We discussed these goals with raumCode during three workshops and, based on this, raumCode was responsible for the technical implementation. Table 1 summarises the capabilities of the social robot and the tasks it carried out during the lecture. A video that shows the tasks can be found here: https://youtu.be/6pRBz hR5kR8.

At the beginning of the lecture, the lecturer pointed out the reasons for using the robot. In large-scale courses it is difficult to offer students an adequate level of support. In such an environment, "Lexi," a Pepper model (see Figure 1), can take on the role of assistant. To this end, it was connected to the projector of the venue and equipped with a headphone. In a face-to-face interaction, Lexi would use its tablet instead of the projector. Overall, Lexi assisted for about 45 minutes. Lexi was permanently listening to the lecturer (signal processing) and answered questions relying on its chatbot. This included explaining something and visualising the content using the projector. However, due to the noise level in the venue, it was not always possible for the speech recognition device to process the signal. This may currently be a problem in the application of social robots (Lehmann & Rossi, 2019). If Lexi was not able to process the signal, a representative from raumCode, who was present during the lecture, manually started the script. At the end of the lecture, the managing director from raumCode explained to the students how the robot works and opened the black box. This included the decisive role of the chatbot software, which enables the robot to adapt to student needs. In our context, this means answering typical questions in the realm

Table 1: Tasks carried out by the social robot

Capability of social robots	*Tasks performed during the "Introduction to academic writing" course*
Greeting	Introduces itself and the institute
Reading	Presents its aim: assisting lecturers and students to foster learning
	Answers the question how plagiarism is detected by means of plagiarism software; explains the software analyses, the outcome of the analytic process and the decision-making process: the human is the final decider about every case (example, for a human–machine interaction and procedure based on complementary skills)
Alerting	Alerts students to use the citation standard of the American Psychological Association
Remote operations	Supports the lecturer by looking for sources about "greenwashing" in the database of the university's library
	Converses with a volunteer student: accesses remote services during the conversation to determine the student's face characteristics, age and mood ("happy," "surprised," "angry," "sad" and "neutral")
Clarification	Presents further material using the projector of the venue and demonstrates the functioning of the plagiarism software by giving an illustrative example
Motions	Follows the lecturer with its head; uses gestures to support its points and to express emotions

Note. Capabilities taken from Cooney and Leister (2019).

of academic writing. Moreover, he explained the process of how the robot detects emotions by means of a Microsoft Azure service. Against this backdrop, students may have obtained a fair picture about what happens behind the scenes.

Method

Sample

We asked all 1552 freshmen via the learning management system to fill in an online questionnaire in English. After 1 week we sent a reminder. After a further week, the survey was closed. This approach yielded a sample of 462 student responses containing no missing values. We only used a questionnaire in English because students from the German track had sufficient command of the English language to fill in the form. Table 2 summarises the characteristics of the sample.

Only 12.77% of the students in the sample had interacted with a social robot like Lexi before attending the "Introduction to academic writing" course. Of the students, 6.93% perceive Lexi as male, 55.41% as female and 37.66% attribute no gender at all to the robot.

Measurement model and moderators

The used constructs, their definition and the items to operationalise them can be found in Table S1 in the Supplementary Files. All items are captured on a seven-point scale of rating, ranging from 1 "entirely disagree" to 7 "entirely agree." SI and appearance are captured using a formative measurement model (Coltman, Devinney, Midgley, & Venaik, 2008). SI may be formed by the influence of lecturers, parents and friends (Seufert, Guggemos, & Sonderegger, 2019). Visual and acoustical stimuli may be responsible for the overall way the robot was perceived because the students saw and heard Lexi, but did not touch (taste or smell) it. Hence, we regard the visual appearance and the voice as constitutive elements for Lexi's appearance (Pandey & Gelin, 2018). Formative measurement instruments have the advantage in that it could be identified to what extent different aspects form the construct, eg, if visual or acoustical impressions are more important in forming a perception about the overall appearance.

Table 2: Characteristics of the sample (N = 462)

Characteristics	*Categories*	*N*	*%*
Gender	Female	163	35.28
	Male	299	64.72
Study track	English	137	29.65
	German	325	70.35
Intended study subject	Business administration	225	48.70
	Economics	102	22.08
	International affairs	71	15.37
	Law	28	6.06
	Law and economics	36	7.79
Native language	German	342	74.02
	French	47	10.17
	Italian	35	7.58
	English	21	4.55
	Others	17	3.68
Age	<19 years	65	14.07
	19 years	158	34.20
	20 years	135	29.22
	21 years	64	13.85
	>21 years	40	8.66

To operationalise the moderating variables, we used students' self-reports: gender, age (20 + and under 20 years), study track, study subject and perceived gender of Lexi. To ascertain previous experience, we asked the question: "Apart from the course 'Introduction to academic writing', have you ever interacted with a social robot like Lexi before?" (0/1).

Data analysis

We use partial least squares structural equation modelling (PLS-SEM: SMART-PLS 3.2.8) to analyse our data. Justification for this is based on the lack of normally distributed data for all items (Shapiro–Wilk test: $p < .05$), our aim to test our model from a prediction perspective (intention to use social robots) and the use of formative measurement models (Hair, Risher, Sarstedt, & Ringle, 2019). Moreover, PLS-SEM is a well-established method in UTAUT research and enables us to identify the most relevant drivers for BI (Ringle & Sarstedt, 2016): importance-performance map analysis (IPMA). In our case, IPMA shows how the nine constructs are shaping BI, which considers both direct and indirect effects (importance). Effects are calculated using unstandardised path coefficients. Students' average latent variable scores on a percentage scale indicate the performance. The goal is to identify those constructs that have a relatively high importance for BI (ie, those that have a strong total effect), but also have a relatively low performance (ie, low average latent variable scores).

PLS-SEM is interpreted in two steps: (i) evaluation of the measurement model and (ii) assessment of the structural model that deals with the relationships between the constructs. To this end, we follow the guidelines of Hair *et al.* (2019): (1) For reflective measurement models we check if all standardised indicator loadings are larger than .708. This may be the prerequisite for a reliable and convergent valid measure. We evaluate reliability using ρ_A (Dijkstra & Henseler, 2015), which may be a good compromise between Cronbach's alpha (too conservative) and composite reliability (too liberal). It should be higher than 0.70 and lower than 0.95. Convergent validity may be ensured if the average variance extracted (AVE) is higher than 0.50 for all constructs; discriminant validity may be established if the heterotrait–monotrait ratio is smaller than 0.90 due to the conceptual similarity of the constructs in the UTAUT. For assessing the quality of formative measurement models, we control for multicollinearity issues by means of the variance inflation factor (VIF). The VIF should be below five for all indicators to avoid problems of multicollinearity. Besides this, we assess if the indicator weights are statistically significant. (2) Concerning the structural model, the VIFs for all paths should be lower than five. Moreover, the in-sample model fit and the out-of-sample predictive power have to be assessed. The determination coefficient R^2 provides evidence for the in-sample explanatory power of the model. Stone–Geisser's Q^2 is a measure for in- and out-of-sample predictive power; values of 0.25 and 0.5 depict medium and large predictive power. For the necessary blindfolding process, we select an omission distance of nine. PLS predict is a relatively new method to assess the out-of-sample predictive power (Shmueli *et al.*, 2019), ie, it could be evaluated if the model is able to predict BI beyond our sample. To this end, training and holdout samples are formed. The parameter estimates from the training sample are used to predict the values for the cases from the holdout sample. First, it is checked whether Q^2_{predict} is larger than zero for all indicators. This would imply a more accurate prediction of our model in comparison to simply taking the means from the training sample. Afterwards, we focus, as suggested (Hair *et al.*, 2019), on BI as the target construct of our model and check if the root mean square error of the PLS-SEM prediction is smaller than a prediction by a linear regression model. If this is the case for two (three) out of the three BI-items, this would imply a medium (high) out-of-sample predictive power.

We conduct multigroup analyses to consider the hypothesised moderators. In this process, we also check for measurement invariance among the groups by means of permutation tests (Henseler, Ringle, & Sarstedt, 2016). Moreover, we carry out a finite-mixture segmentation to control for unobserved heterogeneity. Evidence for the absence of unobserved heterogeneity may be a lower information criterion AIC3 and CAIC for a one-class model in comparison to every multiclass solution (Sarstedt, Becker, Ringle, & Schwaiger, 2011). If the information criteria indicate a multigroup solution, we will perform a predicted-oriented segmentation (Hair *et al*., 2019).

Results

Measurement model

In terms of the reflective measures, we observe standardised factor loadings that exceed .72, ie, at least 51.84% (0.72^2) of the item variance is explained by the construct. The only exception is the second item of "Anxiety handling" with a standardised loading of .67. All ρ_A are above .75, which indicates sufficient reliability; the maximum ρ_A equals .93. Convergent validity is established as the AVE is greater than .68, ie, on average more than 68% of the item variance can be explained by the corresponding construct and less than 32% is error variance. Discriminant validity might be ensured because the heterotrait–monotrait ratio is below .79 with one exception. For PE and EE, it equals .906, which is only slightly above the cut-off value for conceptually similar constructs (0.90). Moreover, the heterotrait–monotrait ratio is significantly lower than one: the 95% confidence interval equals (0.848, 0.965). Hence, the used items to operationalise PE and EE capture different constructs. Table 3 summarises the key characteristics of the measurement model and shows the correlations among the constructs.

In terms of the formative measurement models, all the VIFs are below 2.50, which is ideal. High correlation of indicators that could lead to an imprecise estimation of indicator weights and inflated standard errors are not an issue. All indicator weights are significant ($p < .001$). For "Appearance," we also checked for convergent validity by correlating the formatively measured construct with a reflectively measured one (Item 1: "Lexi is cute"; item 2: "Lexi is boring"). The correlation equals .70, which might be sufficiently high and indicates the suitability of our measurement instrument to capture "Appearance." In sum, we may conclude that our measurement model is sound.

Structural model

The quality of the structural model is decent. Collinearity is not an issue because all the VIFs are lower than 2.43. The explanatory power is moderate with a R^2 of 58.0% for BI ie, 58% of the BI variance can be explained by the structural model. A Stone–Geisser's Q^2 of 0.44 indicates a medium to large predictive accuracy of the model. Moreover, Q^2_{predict} is substantially larger than zero for all indicators. The PLS-SEM root mean square error for the three indicators of BI is lower in comparison to the linear regression model. This may indicate a high out-of-sample predictive power of the model, ie, our model may be useful for predicting BI beyond our sample. Overall, we conclude the soundness of our structural model, which is depicted in Figure 3. Due to the sample size of 462, we are able to detect even small effects ($R^2 = 2.5\%$) at the 5% level with a probability of 80% (statistical power). All direct, indirect and total effects are statistically significant at the 5% level except for those of "Anxiety," ie, all the characteristics of the robot have a significant (indirect) positive association with BI. Table 4 presents the indirect effects including the 95% confidence intervals to assess the statistical significance.

Concerning moderators, we did not find significant differences in path coefficients concerning the perceived gender of Lexi (for all differences: $p > .069$), age group of students (for all path

Table 3: Quality of measurement model and correlation among constructs (N = 462)

				Correlations among constructs; square root of AVE on the diagonal									
	Construct	ρ_A	*AVE*	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
1	PE	0.89	0.90	.95									
2	EE	0.76	0.75	.73	.86								
3	SI	n/a	n/a	.61	.55	n/a							
4	BI	0.92	0.86	.70	.64	.64	.93						
5	Trustworthiness	0.78	0.90	.52	.45	.45	.52	.95					
6	Adaptiveness	0.76	0.80	.66	.58	.46	.56	.50	.90				
7	Appearance	n/a	n/a	.43	.42	.42	.44	.31	.36	n/a			
8	Social presence	0.78	0.68	.40	.42	.49	.48	.43	.30	.39	.82		
9	Anxiety handling	0.79	0.70	−.01	−.10	−.06	−.08	.04	−.03	.02	.05	.84	
10	Anxiety backend	0.93	0.93	−.10	−.11	−.09	−.12	−.14	−.01	−.13	−.11	.12	.96

Note. ρ_A = measure for reliability. AVE = average variance extracted. SI and “Appearance” are formatively measured. Correlations above $|.08|$ are significant at the 5% level, correlations above $|.11|$ are significant at the 1% level.

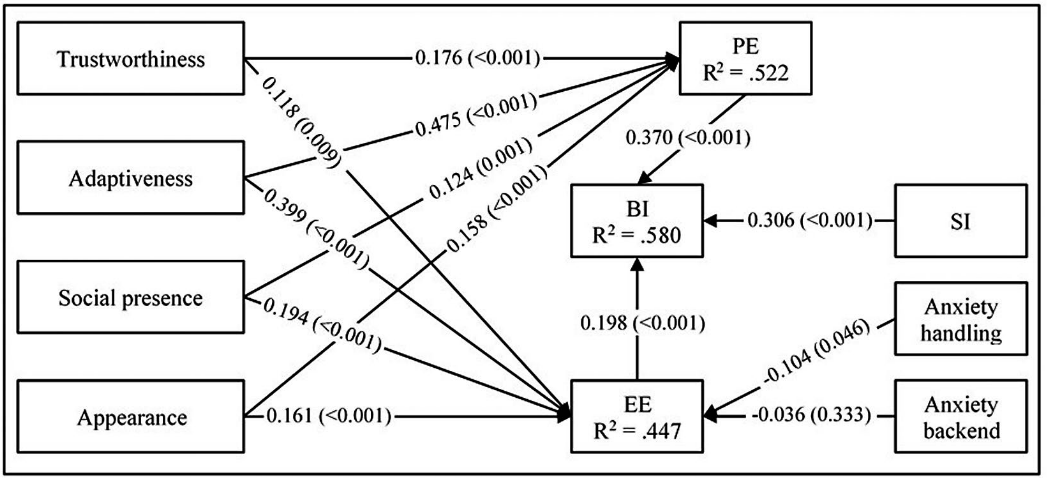


Figure 3: Structural model for predicting Behavioural Intention (N = 462). Note. Standardised path coefficients. Standard errors are calculated by means of bootstrapping (BCa method with 10 000 samples). P-values (two-sided) are depicted in parentheses

differences: $p > .112$) and intended study subject (for all paths differences: $p > .100$). Significantly different path coefficients did occur, however, regarding student gender, study track and previous experience in interacting with social robots. These differences are depicted in Table 5.

The multigroup analyses revealed only three differences in path coefficients (observed heterogeneity). In terms of unobserved heterogeneity, the results are inconclusive. AIC3 indicates a multi-class solution, CAIC (and BIC) a one-class solution. We inspected different class solutions but could not identify groups with meaningful (observable) characteristics. Moreover, the predicted-oriented segmentation did not converge, which might indicate a one-class solution as a suitable model for our data. Against this background and the results of the multigroup analyses, we regard an overarching model for the entire sample as suitable and report IPMA results for the overall sample.

As can be seen from the IPMA depicted in Figure 4, three factors are substantially associated with BI—PE, SI and "Adaptiveness." EE shows a medium predictive power, while "Social presence," "Appearance" and "Trustworthiness" demonstrate a weak predictive power. "Anxiety" does not have a significant (see Table 4) and, in light of the statistical power of our study, practically relevant predictive value for BI. Since PE, the best predictor in the model, is an endogenous variable, we also carried out IPMA for this construct. "Adaptiveness" yields by far the highest importance (0.508), followed by "Trustworthiness," (0.181) "Appearance" (0.176) and "Social presence" (0.144). SI, the second-best predictor for BI, is measured using formative indicators; hence, we can disentangle this variable. The SI of parents has the highest importance (0.144), followed by friends (0.132) and lecturers (0.096).

In terms of performance, there is room for improvement within all constructs. BI shows a performance of 36.6% of the theoretical maximum. Except for two predictors, the performance varies between 40% and 55%. "Anxiety backend" has the highest performance with 64.5%. At the lower end, "Social presence" shows a weak performance with 26.5%.

Table 4: Indirect effects with BI as the outcome variable (N = 462)

	Mediator PE		*Mediator EE*		*Total indirect effect*	
Predictor	*Coef.*	*95%-CI*	*Coef.*	*95%-CI*	*Coef.*	*95%-CI*
Trustworthiness	0.065	(0.033, 0.107)	0.023	(0.006, 0.051)	0.088	(0.046, 0.135)
Adaptiveness	0.176	(0.126, 0.238)	0.079	(0.040, 0.123)	0.255	(0.203, 0.317)
Social presence	0.046	(0.018, 0.079)	0.038	(0.018, 0.068)	0.084	(0.046, 0.127)
Appearance	0.059	(0.029, 0.099)	0.032	(0.012, 0.061)	0.091	(0.050, 0.135)
Anxiety handling	–	–	−0.007	(−0.025, 0.007)	−0.007	(−0.025, 0.007)
Anxiety backend	–	–	−0.021	(−0.045, 0.006)	−0.021	(−0.045, 0.006)

Note. Confidence intervals (CI) calculated by means of bootstrapping (BCa method with 10 000 samples). Coef. = standardised path coefficient.

Table 5: Multigroup analyses: Significantly different path coefficients

Path	*Moderator*	*Manifestation*	*Coef.*	*|Dif.|*	*p-value*
Anxiety backend → EE	Gender	Female ($n = 163$)	−0.171	0.198	.003
		Male ($n = 299$)	0.027		
Social presence → EE	Study track	German ($n = 325$)	0.250	0.153	.025
		English ($n = 137$)	0.097		
Appearance → PE	Robot experience	Yes ($n = 59$)	−0.049	0.242	.009
		No ($n = 403$)	0.194		

Coef. = standardised path coefficient. Dif. = difference in path coefficients.

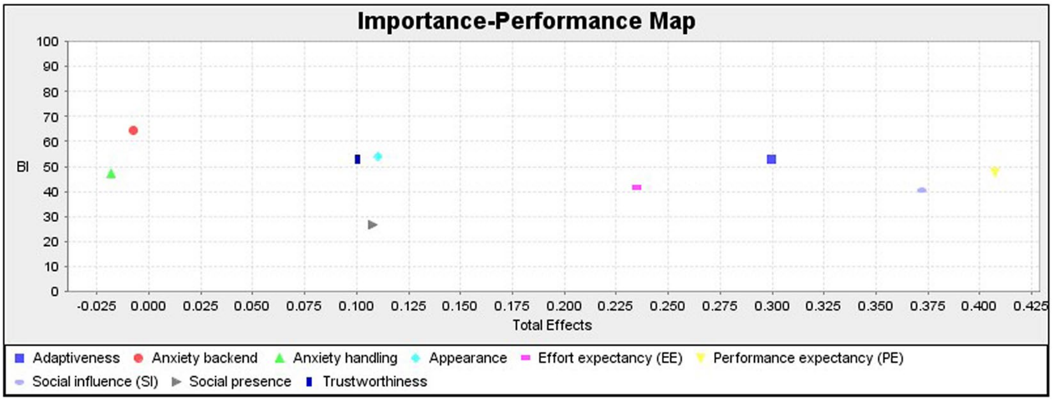


Figure 4: Importance-performance map analysis (IPMA) for Behavioural Intention (BI)

Discussion

Findings

Offering adequate support to students in large-scale higher education courses is often challenging (Handke, 2018); relying on social robots could be a viable option to tackle this issue (Byrne, Rossi, & Doolan, 2017; Cooney & Leister, 2019). The context for our study is an academic writing course that may be representative of many large-scale higher education courses; the heterogeneity, especially in terms of prior knowledge about academic writing, is high (Seufert & Spiroudis, 2017). Since academic writing follows rules, eg, codified in the Publication Manual of the American Psychological Association (APA, 2020), a social robot could assist the lecturer by answering typical student questions. However, during the process of deciding whether or not to introduce social robots into higher education courses, their acceptance by students has to be evaluated. This leads to the first research question. Concerning BI and its predictors (RQ1), the students in our sample overall do not have the intention to rely on social robots for learning purposes, with BI reaching only 36.6% of the theoretical maximum.

One reason could be that the lecturer, rather than the students, interacted with the robot. The assessment of social robots depends on the level of engagement with them (Conti *et al.*, 2019; Cooney & Leister, 2019; Fridin & Belokpytov, 2014). A face-to-face interaction over a substantial period of time is likely to be more intense than the setting in our study, where the lecturer interacted with the robot for about 45 minutes. Based on a more intense interaction, the assessment of the social robot could be more elaborate. However, we controlled for prior experience with social robots by asking students if they had interacted with a social robot like Lexi apart from in the course "Introduction to academic writing," eg, in a shopping mall. The results show that students do not differ in attitude depending on their previous experience; their BI performance equals 36.7%, as opposed to 36.6%. Hence, prior experience with social robots may not have a substantial influence on BI.

On a conceptual level, the UTAUT implies that the constructs PE, SI and EE are responsible for a low BI. Of these, PE and SI showed the highest importance with values of 0.407 and 0.372, respectively. Following the idea of IPMA, focussing on PE in order to increase BI may be advantageous because it has the highest importance and offers substantial room for improvement (performance = 47.7%). PE is conceptualised as the degree to which students believe that using social robots will be helpful in their learning process. Drawing on research about high-quality learning settings (from a student perspective), three facets are relevant (Praetorius *et al.*, 2014): classroom management, learning support and cognitive activation. Since our demonstration of the robot's capabilities did not include activities of classroom management, the latter two may be relevant for student assessment of Lexi. Learning support involves all activities that aim at motivating and engaging students in learning. Lexi answered the questions at the factual level. Activities that aimed at establishing a personal relationship between the lecturer and Lexi, eg, using the lecturer's name, were limited. Moreover, Lexi did not provide feedback, eg, "well-done." Based on our knowledge regarding typical questions in the course, Lexi answered the questions in a way potentially suitable for a broad variety of students. However, these answers were not tailored to individual needs.

Cognitive activation comprises "providing challenging tasks in zones of proximal development, activating previous knowledge, building on students' ideas and experiences, and posing stimulating questions" (Praetorius *et al.*, 2014, p. 3). As can be seen from the description of the tasks in Table 1, Lexi did not offer tasks. Rather, it answered the questions of the lecturer without considering her prior knowledge. Furthermore, Lexi did not raise activating questions. Overall, when comparing the activities carried out by Lexi, shown in Table 1, with the criteria for a high-quality learning setting, the finding of a moderate PE is not surprising. This also implies that at the current level of state-of-the-art technology there is much room for improvement in the capabilities of the social robot to help establish a high-quality learning setting. This might be consistent with the findings of Cheng, Sun and Chen (2018).

Social influence, at 0.372, has a high importance similar to that of PE. Moreover, with a performance of 40.5% there is substantial room for improvement. Hence, when introducing social robots into higher education, the perceptions of parents, friends and lecturers should be taken into account. To increase the general acceptance of social robots as learning assistants, a communication strategy may be necessary. For our part, we have created a video that introduces Lexi and its capabilities and disseminated it via our social media channels and webpage. Moreover, the local press reported the use of Lexi as a teaching assistant. Such communication activities and media coverage can increase SI (Alaiad & Zhou, 2014).

Social robot research has identified characteristics that may be important for their acceptance by users (eg, Graaf & Allouch, 2013). Concerning the characteristics of the robot that predict BI (RQ2), "Adaptiveness" shows by far the highest predictive power. Moreover, there is room for improvement (performance = 53.1%). These findings may be consistent with established theory about learning and teaching. According to Bransford, Brown and Cocking (2000), environments conducive to learning should be learner centred. This means student background and prior knowledge is considered in order to offer learning opportunities at a suitable level of difficulty.

High-quality learning environments are also assessment centred through offering continuous feedback (Bransford, Brown, & Cocking, 2000). The positive impact of continuous feedback on learning outcomes is also true for higher education (Hattie, 2015). Providing feedback may be regarded as a manifestation of adaptiveness; however, adaptiveness might be the most difficult characteristic for the robot to increase. It heavily depends on artificial intelligence in order to consider students' prior knowledge, provide learning material at a suitable level of difficulty, select an appropriate learning pace and give individualised feedback. In light of this, educational psychology, research about social robotics and artificial intelligence may go hand in hand. An example of this could be to use pupil dilation to measure mental effort (van der Wel & van Steenbergen, 2018). Based on face recognition (artificial intelligence), the pupil diameter of the student could be traced and the difficulty of learning material adjusted accordingly to ensure a suitable cognitive load. A further example concerns emotions, which play a crucial role in learning (Tyng, Amin, Saad, & Malik, 2017). These could be detected based on facial expressions (artificial intelligence) and utilised to adjust instructions accordingly. Overall, it seems to be important to have a strong conceptual basis of high-quality learning settings and, based on this, decide how artificial intelligence can offer solutions that social robots can then carry out.

In comparison to "Adaptiveness," other robot characteristics are not as important. However, if "Appearance" (importance = 0.111) were to be improved, voice is (statistically) significantly more important than visual appearance, 0.072 compared with 0.039. "Social presence," ie, the capability of a robot to show "emotions" and act like a real person, has a similar importance to "Appearance" (0.108). Hence, despite the current quite low performance in this area (26.5%), it may not be a priority to focus on "Social presence" in order to increase BI. This may also be true for "Trustworthiness," which might already be at an acceptable level (53.0%).

Concerning anxiety (RQ3), surprisingly, "Anxiety backend" does not predict BI, although it is the construct with the highest performance (65.0%), which indicates serious concerns about data protection and privacy. We also controlled for a direct association with BI; it was not significant (−0.031, $p = .212$). These findings might be consistent with the privacy paradox (Acquisti, Brandimarte, & Loewenstein, 2015): people report serious concerns about privacy issues, but voluntarily reveal private information and continue using services they reportedly distrust. Lutz and Tamó-Larrieux (2020) provided empirical evidence for this phenomenon in the context of social robots. In line with our findings, they reported a non-significant association between privacy concerns and BI. Against this backdrop, it may not be necessary to mitigate students' privacy concerns in order to increase BI as they are not a significant predictor. However, transparency regarding what happens behind the scenes of a social robot may be required from an ethical and legal perspective (Felzmann, Fosch-Villaronga, Lutz, & Tamo-Larrieux, 2019). Moreover, in the medium or long-term, privacy concerns may undermine the trustworthiness of social robots or their perception by the public, which in turn could be detrimental to SI. For our part, the managing director of raumCode explained to the students how the robot works and which services it uses in the backend. Hence, we enabled students to make an informed assessment of the social robot.

Limitations

Our study is not without limitations. Based on the fact the demonstration was only 45 minutes long, students might not have had enough time to obtain a comprehensive picture of the robot's capabilities. Moreover, we have only investigated student intention to use social robots for learning purposes and its predictors. As such, it is not possible to determine whether the use of social robots leads to an increase in knowledge or skills. Student perceptions about what is conducive to learning can be different from research results (Kirschner & van Merriënboer, 2013). The constructs explored, using our items, are narrow in scope and their full complexity is not covered, eg, in the case of trust. Moreover, we cannot tell how social robots perform in comparison to other forms of technology-based assistance. Evidence in this regard would be necessary to evaluate if the additional expenses incurred by a social robot, in comparison to a virtual agent, are justified.

Future work

Overall, 58% of the variance regarding the intention to use social robots can be explained. Further research could aim at identifying additional variables to explain a higher proportion of variance. Hedonic motivation could be such a predictor (Graaf & Allouch, 2013; Venkatesh, Thong, & Xu, 2012). Moreover, the kind of facilitating conditions important for students in higher education could be investigated in order to predict the actual use of social robots. Due to the high importance of adaptiveness, it might be promising to split up this predictor by working out what constitutes adaptiveness based on theory (eg, Brühwiler & Blatchford, 2011). Facets could include considering students' background or providing individualised support. A corresponding formative measurement instrument could then be used to identify which facets are more, or less, important for forming the adaptiveness of a social robot.

Adaptiveness might be necessary to ensure an assessment-centred learning environment. Community centredness is a further element in a high-quality learning environment (Bransford, Brown, & Cocking, 2000), where learners support each other. This may also be true for academic writing (Guasch, Espasa, Alvarez, & Kirschner, 2013). In this vein, the social presence of the robot might play an important role (Howley, Kanda, Kotaro, & Rosé, 2014). How social robots and humans can effectively collaborate to create a community-centred learning environment might be a promising avenue for future research.

Conclusion

The aim of this study was to investigate the acceptance of social robots for learning purposes by students from the social sciences. A Pepper-type model from SoftBank Robotics provided a sample of its capabilities during an introductory university course. The UTAUT acted as the theoretical background. All variables implied by the UTAUT—PE, EE and SI—positively predict BI. The perceived characteristics of the social robot—trustworthiness, adaptiveness, social presence and appearance—indirectly predict BI mediated by PE and EE. Surprisingly, concerns about privacy issues do not negatively predict BI. Multigroup analyses and finite-mixture segmentations may indicate that our results are robust in terms of observed and unobserved heterogeneity. If BI were to be increased, the focus could be on adaptiveness because this characteristic shows the highest importance and, at the same time, a moderate performance, ie, the association with BI is substantially higher in comparison to all other perceived characteristics and there is room for improvement.

Statements on open data, ethics and conflict of interest

Readers who wish to examine the data can contact the first author.

The procedures performed in this study follow the standards of the Ethics Committee of the University of St.Gallen.

The authors have no conflict of interest.

References

Abildgaard, J. R., & Scharfe, H. (2012). A geminoid as lecturer. In S. S. Ge, O. Khatib, J. J. Cabibihan, R. Simmons, & M. A. Williams (Eds.), *Social Robotics—4th International Conference, ICSR 2012, Proceedings* (pp. 408–417). Chengdu, China. https://doi.org/10.1007/978-3-642-34103-8_41

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Alaiad, A., & Zhou, L. (2014). The determinants of home healthcare robots adoption: An empirical investigation. *International Journal of Medical Informatics*, *83*(11), 825–840. https://doi.org/10.1016/j.ijmedinf.2014.07.003

APA. (2020). *Publication manual of the American Psychological Association: The official guide to APA style* (7th ed.). Washington, DC: Author.

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, *102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Bacivarov, I. C., & Ilian, V. L. M. (2012). The paradigm of utilizing robots in the teaching process: a comparative study. *International Journal of Technology and Design Education*, *22*(4), 531–540. https://doi.org/10.1007/s10798-011-9157-5

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), 1–9. https://doi.org/10.1126/scirobotics.aat5954

Bolea, Y., Grau, A., & Sanfeliu, A. (2016). From research to teaching: Integrating social robotics in engineering degrees. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *10*(6), 1020–1023. https://doi.org/10.5281/zenodo.1124667

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, *21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Byrne, B. M., Rossi, A., & Doolan, M. (2017). Humanoid robots attending lectures. In *Proceedings of the 9th International Conference on Education Technology and Computers, ICETC*, Barcelona, Spain. https://doi.org/10.1145/3175536.3175556

Cheng, Y.-W., Sun, P.-C., & Chen, N.-S. (2018). The essential applications of educational robot: Requirement analysis from the perspectives of experts, researchers and instructors. *Computers & Education*, *126*, 399–416. https://doi.org/10.1016/j.compedu.2018.07.020

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262. https://doi.org/10.1016/j.jbusres.2008.01.013

Conti, D., Cattani, A., Di Nuovo, S., & Di Nuovo, A. (2019). Are future psychologists willing to accept and use a humanoid robot in their practice? Italian and English students' perspective. *Frontiers in Psychology*, *10*, 1–13. https://doi.org/10.3389/fpsyg.2019.02138

Cooney, M., & Leister, W. (2019). Using the engagement profile to design an engaging robotic teaching assistant for students. *Robotics*, *8*(1), 1–26. https://doi.org/10.3390/robotics8010021

de Graaf, M., & Allouch, S. B. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, *61*(12), 1476–1486. https://doi.org/10.1016/j.robot.2013.07.007

Devitt, S. K. (2018). Trustworthiness of autonomous systems. In H. A. Abbass, J. Scholz, & D. J. Reid (Eds.), *Foundations of trusted autonomy* (pp. 161–84). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-64816-3_9

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, *39*(2), 297–316. https://doi.org/10.25300/MISQ/2015/39.2.02

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Robots and transparency: The multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics & Automation Magazine*, *26*(2), 71–78. https://doi.org/10.1109/MRA.2019.2904644

Fridin, M., & Belokpytov, M. (2014). Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Computers in Human Behavior*, *33*, 23–31. https://doi.org/10.1016/j.chb.2013.12.016

Gao, Y., Barendregt, W., Obaid, M., & Castellano, G. (2018). When robot personalisation does not help: insights from a robot-supported learning study. In *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 705–712). Nanjing: IEEE. https://doi.org/10.1109/ROMAN.2018.8525832

Gardecki, A., & Podpora, M. (2017). *Experience from the operation of the Pepper humanoid robots*. Piscataway, NJ: IEEE. https://doi.org/10.1109/PAEE.2017.8008994

Garone, A., Pynoo, B., Tondeur, J., Cocquyt, C., Vanslambrouck, S., Bruggeman, B., & Struyven, K. (2019). Clustering university teaching staff through UTAUT: Implications for the acceptance of a new learning management system. *British Journal of Educational Technology*, *50*(5), 2466–2483. https://doi.org/10.1111/bjet.12867

Guasch, T., Espasa, A., Alvarez, I. M., & Kirschner, P. A. (2013). Effects of feedback on collaborative writing in an online learning environment. *Distance Education*, *34*(3), 324–338. https://doi.org/10.1080/01587919.2013.835772

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Handke, J. (2018). Digitale Hochschullehre—Vom einfachen Integrationsmodell zur Künstlichen Intelligenz [Digital higher education teaching - from a simple integration model to artificial intelligence]. In U. Dittler, & C. Kreidl (Eds.), *Hochschule der Zukunft [Higher education of the future]* (pp. 249–263). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-20403-7_15

Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79–91. https://doi.org/10.1037/stl0000021

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: The Almere model. *International Journal of Social Robotics*, *2*(4), 361–375. https://doi.org/10.1007/s12369-010-0068-5

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, *33*(3), 405–431. https://doi.org/10.1108/IMR-09-2014-0304

Howley, I., Kanda, T., Hayashi, K., & Rosé, C. (2014). Effects of social presence and social role on help-seeking and learning. In G. Sagerer, M. Imai, T. Belpaeme, & A. Thomaz (Eds.), *HRI'14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 415–422). New York, NY: ACM Press. https://doi.org/10.1145/2559636.2559667

Huang, C.-M., & Mutlu, B. (2014). Learning-based modeling of multimodal behaviors for humanlike robots. In G. Sagerer, M. Imai, T. Belpaeme, & A. Thomaz (Eds.), *HRI'14: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 57–64). New York, NY: ACM Press. https://doi.org/10.1145/2559636.2559668

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). The robot who tried too hard. In J. A. Adams, W. Smart, B. Mutlu, & L. Takayama (Eds.), *Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 67–74). New York, NY: ACM. https://doi.org/10.1145/2696454.2696457

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? *Urban Legends in Education. Educational Psychologist*, *48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Lee, J.-H., & Song, C.-H. (2013). Effects of trust and perceived risk on user acceptance of a new technology service. *Social Behavior and Personality: An International Journal*, *41*(4), 587–597. https://doi.org/10.2224/sbp.2013.41.4.587

Lehmann, H., & Rossi, P. G. (2019). Social robots in educational contexts: Developing an application in enactive didactics. *Journal of e-Learning and Knowledge Society*, *15*(2), 27–41. https://doi.org/10.20368/1971-8829/1633

Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media & Communication*, *7*(3), 412–434. https://doi.org/10.1177/2050157919843961

Lutz, C., & Tamó-Larrieux, A. (2020). The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots. *Human-Machine Communication*, *1*, 87–111. https://doi.org/10.30658/hmc.1.6

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Technology for Education and Learning*, *1*(1), 1–7. https://doi.org/10.2316/Journal.209.2013.1.209-0015

Nistor, N., Stanciu, D., Lerche, T., & Kiel, E. (2019). "I am fine with any technology, as long as it doesn't make trouble, so that I can concentrate on my study": A case study of university students' attitude strength related to educational technology acceptance. *British Journal of Educational Technology*, *50*(5), 2557–2571. https://doi.org/10.1111/bjet.12832

Pandey, A. K., & Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, *25*(3), 40–48. https://doi.org/10.1109/MRA.2018.2833157

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, *31*, 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002

Ringle, C. M., & Sarstedt, M. (2016). Gain more insight from your PLS-SEM results. *Industrial Management & Data Systems*, *116*(9), 1865–1886. https://doi.org/10.1108/IMDS-10-2015-0449

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with Fimix-Pls: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review*, *63*, 34–62. https://doi.org/10.1007/BF03396886

Seufert, S., Guggemos, J., & Sonderegger, S. (2019). Learning analytics in higher education using peer-feedback and self-assessment: Use case of an academic writing course. In *Proceedings of the 11th International Conference on Computer Supported Education* (pp. 315–322). Heraklion: CSEDU. https://doi.org/10.5220/0007714603150322

Seufert, S., & Spiroudis, E. (2017). Wissenschaftliche Studieneingangsphase fördern: Erfahrungen mit einem Rückmelde- und Bewertungsraster (Rubric) [Support the scientific study entry phase: Experiences with a feedback and evaluation grid (rubric)]. *Zeitschrift Schreiben*, *15*, 1–9. Retrieved from https://zeitschrift-schreiben.eu/globalassets/zeitschrift-schreiben.eu/2017/seufert_spiroudis_studieneingangsphase.pdf

Shin, D. H., & Choo, H. (2011). Modeling the acceptance of socially interactive robotics: Social presence in human–robot interaction. *Interaction Studies*, *12*(3), 430–460. https://doi.org/10.1075/is.12.3.04shi

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Spolaôr, N., & Benitti, F. B. V. (2017). Robotics applications grounded in learning theories on tertiary education: A systematic review. *Computers & Education*, *112*, 97–107. https://doi.org/10.1016/j.compedu.2017.05.001

Tanaka, F., Isshiki, K., Takahashi, F., Uekusa, M., Sei, R., & Hayashi, K. (2015). Pepper learns together with children: Development of an educational application. In T. Asfour (Ed.), *15th International Conference on Humanoid Robots (Humanoids), IEEE-RAS: November 3–5, 2015, Seoul, Korea* (pp. 270–275). Piscataway, NJ: IEEE. https://doi.org/10.1109/HUMANOIDS.2015.7363546

Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics. *IEEE Robotics & Automation Magazine*, *14*(1), 35–42.

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., & Koay, K. L. (2008). Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, *24*(2), 159–178. https://doi.org/10.1007/s10514-007-9058-3

Zhong, B., &Xia, L. (2018). A systematic review on exploring the potential of educational robotics in mathematics education. *International Journal of Science and Mathematics Education*, *18*, 79–101. https://doi.org/10.1007/s10763-018-09939-y

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Studie 6:

Guggemos, J., Sonderegger, S. & Seufert, S. (2022). Student acceptance of social robots in higher education – Evidence from a vignette study. In *Proceedings of the 19th International Conference on Cognition and Exploratory Learning in Digital Age,* 105–114.

STUDENT ACCEPTANCE OF SOCIAL ROBOTS IN HIGHER EDUCATION – EVIDENCE FROM A VIGNETTE STUDY

Josef Guggemos[1], Stefan Sonderegger[2] and Sabine Seufert[2]
[1]*University of Education Schwäbisch Gmünd, Hauberweg 48, 73525 Schwäbisch Gmünd, Germany*
[2]*University of St.Gallen, St. Jakob-Strasse 21, 9000 St. Gallen, Switzerland*

ABSTRACT

Social robots have the potential to play a vital role in education. Student technology acceptance may be considered as a driver for their successful integration in educational processes. Based on the ICAP (Interactive, Constructive, Active, Passive) framework and research on social robot roles, we have developed eight vignettes describing scenarios for the use of social robots in higher education, e.g., teaching assistant. Students in an introductory university course completed the vignettes; they answered questions based on the unified theory of acceptance and use of technology (UTAUT), as well as on their ethical approval. Drawing from a sample of N = 361 students, we carried out confirmatory factor analysis. In line with the conceptual basis, eight scenarios can be described. Overall, students do not accept social robots as presenters; however, the assessment of other roles, such as teaching assistant, is positive. By means of a latent profile analysis, we were able to identify four profiles that may be meaningful from a conceptual point of view. One profile is particularly noteworthy. It comprises students who accept social robots not as a means of instruction, e.g., as a tutor, but as a tool, e.g., to promote computational thinking.

KEYWORDS

Social Robot, Higher Education, Technology Acceptance, Vignette, ICAP, Robot Roles

1. INTRODUCTION

Social robots have the potential to enrich educational processes (Belpaeme et al., 2018; Woo et al., 2021). A social robot can be defined as "an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact" (Bartneck & Forlizzi, 2004, p. 592). Social robots can be used to teach about the robots themselves or as teaching aids (Mubin et al., 2013). When teaching about social robots, they are the actual content of instruction, for example, to teach computational thinking (Ching et al., 2018). Besides this, social robots (in collaboration with teachers) can carry out selected duties in the classroom or venue. Such roles are: teacher, teaching assistant, evaluator, tutor, and peer (Woo et al., 2021). If social robots should take over these roles, however, investigating technology acceptance may be an important consideration (Raffaghelli et al., 2022). Regardless of their actual value, social robots might only be useful if students are willing to use them.

The paper at hand focuses on the use of social robots in higher education, in particular, on academic writing. This domain is characterised by substantial student heterogeneity (Seufert & Spiroudis, 2017). Therefore, a general issue in higher education may be particularly problematic in this context: it is difficult to offer students sufficient guidance and support (Byrne et al., 2017; Cooney & Leister, 2019; Handke, 2018). Employing human lecturers for this purpose may be impossible, mainly due to budget restrictions. Social robots might be a viable option. Studies about the general acceptance of social robots by students for learning purposes are available in this context (Guggemos et al., 2020). However, these studies in general address the overall technology acceptance of students in higher education. This may not be satisfactory because the roles that a social robot can adopt are manifold (Belpaeme & Tanaka, 2021; Mubin et al., 2013; Woo et al., 2021). A first important question would be whether the roles of social robots as pointed out in the literature are meaningful, from a pedagogical point of view, and can they be empirically separated. Gauging student acceptance for specific robot roles may be beneficial for developing further use cases. If students, for instance, do not accept a social robot as a presenter but do so as a tutor, further developments might focus on the use of such robots in a tutor setting, instead of in a presenter setting. Besides this, there may not only be variance in the assessment of social robot roles; differences within specific roles may also be present. For instance, some students may accept social robots as a tutor whereas other students, for a variety of reasons, do not accept them for this purpose. Identifying the profiles of students with a similar acceptance of social robots allows for person-centered interventions (Hofmans et al., 2020). For example, the concerns of students who are overly skeptical about using social robots could be addressed. The paper at hand aims to contribute to a more detailed picture of technology acceptance by higher education students in comparison to the available studies on the subject.

2. THEORETICAL BACKGROUND

A first step is to classify meaningful pedagogical scenarios of social robot use. To this end, we rely on the ICAP framework (Chi & Wylie, 2014). It can be used to predict the level of learning outcomes of different activities. The acronym ICAP stands for the learning activities *Interactive*, *Constructive*, *Active*, and *Passive*. The authors argue and provide empirical evidence that in terms of learning outcomes it can be expected that: Interactive > Constructive > Active > Passive. In passive learning activities, learners absorb information from the material without engaging in learning-related activities, e.g., listening to a social robot without further activities. Active learning activities are characterized by activities that do not go beyond the information presented, e.g., making verbatim notes while listening to the robot. In constructive learning activities, learners generate output that goes beyond the information contained in the material, e.g., using the robot as a tool to perform computational thinking tasks. In interactive activities, learners perform the constructive activities with others (including a robot), e.g., solving a problem in collaboration with the social robot.

The interaction with a social robot can be one-to-one (human to robot) or between a robot and multiple humans (multi-party interaction) (e.g., Żarkowski, 2019). This classification is particularly relevant from a technical perspective. The requirements for a multi-party interaction involving the perception of various social cues and linguistic input from multiple individuals is substantially more complex than for a human-to-robot interaction (e.g., identifying who is the sender and who is the receiver of a speech input). Belpaeme et al. (2018) adopted this classification in their meta-analysis. In the studies under consideration, the robot is used to teach one learner in around two-thirds of the cases, and to teach many learners in around one-third of the cases. Woo et al. (2021), who address social robots in classroom settings, distinguish between whole-class instruction and one-to-one interactions.

An established theory to evaluate student acceptance of social robots is the unified theory of acceptance and use of technology (UTAUT) (Fridin & Belokopytov, 2014). The UTAUT (Venkatesh et al., 2003) implies that the use intention can be predicted with performance expectancy, effort expectancy, and social influence. Performance and effort expectancy could be classified as internal beliefs, and social influence as external beliefs (Scherer et al., 2020). Adapted to our context, these constructs are defined as: the degree to which students believe that using social robots in a specific role will benefit them (performance expectancy), the degree of ease associated with the use of social robots in a specific role (effort expectancy), the degree to which students believe others want them to use social robots in a specific role (social influence), and the intention to use the social robot in a specific role (behavioral intention) (Venkatesh et al., 2003). Besides these generic constructs, technology, such as social robots, also brings with it (new) moral challenges (Serholt et al., 2017; Sharkey, 2016; Smakman et al., 2021). Hence, it may also be important to consider student ethical approval about the use of social robots in order to obtain a full picture of technology acceptance. Ethical approval could be defined as the degree to which students regard the use of social robots in a specific role as morally correct.

3. THE PRESENT STUDY

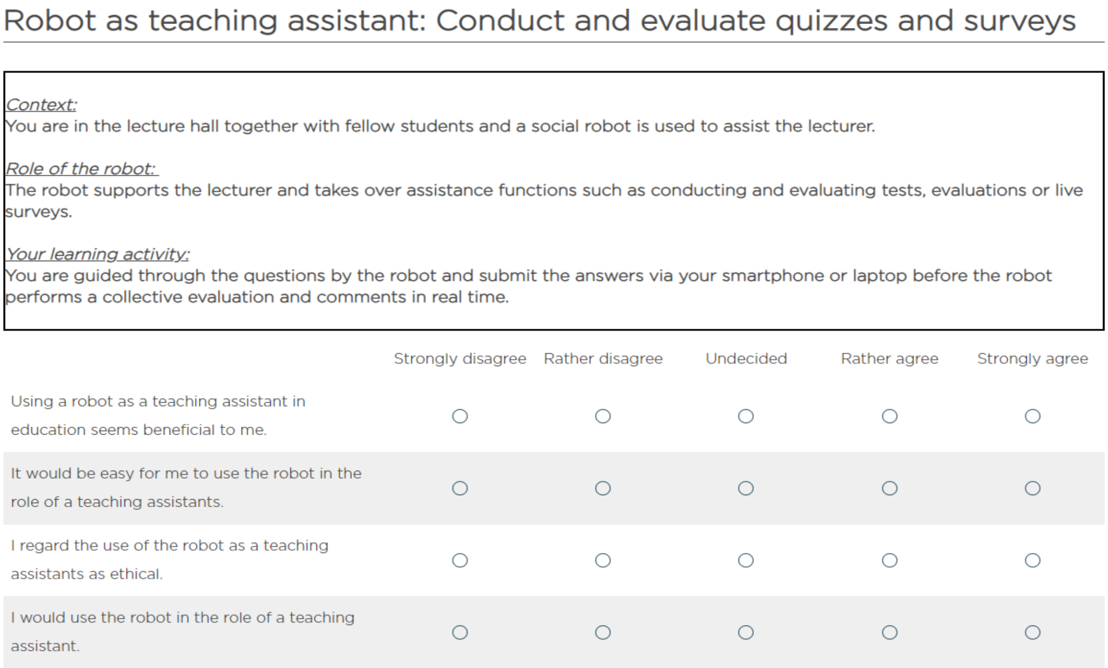
Social robots are a novel technology for higher education students, especially in the social sciences (Guggemos et al., 2020). Hence, it might be difficult for such students to assess the various roles social robots can play. A viable option to tackle this issue may be vignettes (Sailer et al., 2021). Vignettes are stimuli that present realistic scenarios in order to evoke a reaction from the study participants (Skilling & Stylianides, 2020). It may be advantageous to combine both text and videos. Hence, before performing the vignettes, all students watched a video demonstrating the capabilities of social robots, capabilities relevant for all the vignettes (https://youtu.be/GTyKb8c2b6w). The social robot in the video is a Pepper model from Softbank robotics. As previously demonstrated, Pepper may be suitable for educational purposes (Alnajjar et al., 2021) and the robot has been increasingly used in education in recent years (Woo et al., 2021). Drawing from the theoretical background, we developed four vignettes according to the ICAP framework where the robot acts in a robot-to-class setting, and four vignettes in a one-student setting. Following Skilling and Stylianides (2020), the written part of each vignette ranges between 50 and 200 words. Table 1 provides an overview of the eight vignettes.

We validated the eight vignettes in an expert interview with Prof. Handke, a renowned researcher in the realm of social robots in higher education (e.g., Alnajjar et al., 2021), and we revised the eight vignettes based on his feedback. By means of this, we may have obtained a fair representation of social robots' capabilities in higher education (see Table 1).

Table 1. Summary of eight vignettes presenting robot scenarios in higher education

Vignette	Summary of scenario	Added value of the robot	Conceptual basis
Robot-to-class scenarios (1-many)			
1. Presenter (pre) *(passive learning)*	The robot assumes the role of a content presenter in the classical teaching format for knowledge transfer, while the students have a passive role.	Relief for lecturers; variety for students; efficiency benefits; monitoring functions (e.g., attention, volume, air)	~Role of presenter (Handke, 2020, p. 111)
2. Teaching assistant (tas) *(active learning)*	The robot supports the lecturer in a teaching assistant role by conducting and evaluating quizzes or live surveys that students submit.	Support for lecturers; activation of students; evaluations in real time through connection to cloud-based survey tools or LMS	~Role of assistant (Handke, 2020, p. 122)
3. Formative assessor (fas) *(constructive learning)*	The robot, in the role of an examiner, conducts and analyzes competence-oriented assessment tasks, which are solved individually before evaluation and discussion in a plenum.	Relief for lecturers in conducting assessments; consistent and fair exam preparation across courses or years; AI can review text-based answers faster than humans	~Role of examiner (Handke, 2020, p. 127)
4. Moderator (mod) *(interactive learning)*	The robot takes over the moderation and organization of discussions, debates, or group work. Students discuss ideas and lines of argumentation under the robot's guidance.	A robot moderator can store, summarize, and analyze discussion content with AI-based services (e.g., cluster analysis, subjectivity, fact check)	~Role of moderator (e.g., Short et al., 2016)
Robot-to-student scenarios (1-1)			
5. Tutor (tut) *(passive learning)*	The robot adopts the role of a personal tutor offering adaptive repetition and tutoring of content during and after courses to students.	Individual support for students; content is visually/auditorily enhanced and patiently repeated on demand (efficiency)	~Role of tutor (Handke, 2020, p. 106)
6. Advisor (adv) *(active learning)*	The robot, as a learning advisor, answers questions, conducts self-assessments, and provides learning guidance based on learning data.	Recurring admin. questions answered by robot to relieve staff; robot advisory functions based on learning analytics (LMS data)	~Role of advisor (Handke, 2020, p. 118)
7. Tool (tol) *(constructive learning)*	The robot is used as a hands-on learning tool to promote problem solving skills, programming, and computational thinking.	Hands-on orientation provides higher forms of engagement (constructive learning); project-based learning	~Robot as a tool (Handke, 2020, p. 141)
8. Conversation partner (cov) *(interactive learning)*	The robot, as a conversation partner, holds and simulates conversations with students and gives feedback (e.g., simulation of job interview).	Scalable solution to simulate human conversation and learning partners; 'learning by teaching' effect for students	~Conversation practice (e.g., Engwall et al., 2021)

Figure 1 depicts one of the eight vignettes. As can be seen, four questions are embedded in each vignette (in English). Three of these questions are based on the key technology acceptance variables of performance expectancy, effort expectancy, and use intention. We have omitted social influence as it is an external belief. The self-determination theory indicates that such external beliefs could be problematic (Deci et al., 2017); if prompted by others, students may use social robots but with low engagement. We complemented these technology acceptance questions with one question concerning the ethical use of robots.


Robot as teaching assistant: Conduct and evaluate quizzes and surveys

Context:
You are in the lecture hall together with fellow students and a social robot is used to assist the lecturer.

Role of the robot:
The robot supports the lecturer and takes over assistance functions such as conducting and evaluating tests, evaluations or live surveys.

Your learning activity:
You are guided through the questions by the robot and submit the answers via your smartphone or laptop before the robot performs a collective evaluation and comments in real time.

	Strongly disagree	Rather disagree	Undecided	Rather agree	Strongly agree
Using a robot as a teaching assistant in education seems beneficial to me.	○	○	○	○	○
It would be easy for me to use the robot in the role of a teaching assistants.	○	○	○	○	○
I regard the use of the robot as a teaching assistants as ethical.	○	○	○	○	○
I would use the robot in the role of a teaching assistant.	○	○	○	○	○

Figure 1. Vignette example – Robot as teaching assistant

Based on the student assessment of the eight scenarios, we aim to answer two research questions:
RQ1: What is the factor structure of the eight scenarios of social robot use in higher education?
RQ2: What latent profiles of social robot acceptance exist among students in higher education?

4. METHOD

4.1 Sample

First-year students from the course ‘Introduction to Academic Writing’ (about 1,500 participants) at the University of St.Gallen in Switzerland were asked to perform the eight vignettes in November 2021. Of this group, 371 students voluntarily participated. We checked for multivariate outliers by means of Mahalanobis distances. We excluded ten students with significant (α = .01) Mahalanobis distances, which yields a sample of 361 students. Of the data, 0.3% is missing; we inspected the data and concluded it was missing completely at random. Of the students, 36.7% are female and 63.3% male. Their intended study subjects are business administration (43.4%), economics (15.0%), international affairs (11.4%), law (6.1%), and law and economics (8.3%); 15.8% of the students are undecided. On average, the students are 19.71 years old (SD = 1.52 years), min = 17 years, max = 27 years, median = 20 years. The German language track was selected by 63.2% of the students, and the remainder (36.8%) chose the English language track.

4.2 Confirmatory Factor Analysis, Model Assessment, Factor Extraction (RQ1)

To answer RQ1, we carry out confirmatory factor analysis using the ‘lavaan 0.6-11’ package in R (Rosseel, 2012). Since all the items are measured on a five-point scale of rating, we utilize a robust maximum likelihood estimator (MLR) (Robitzsch, 2020). Missing data are handled using full information maximum likelihood estimation. According to Hu and Bentler (1999), a CFI and TLI > 0.95, as well as a RMSEA < 0.06 and SRMR < 0.08, may indicate a good fit of a model. In addition to demonstrating a sufficient fit of a hypothesized model, it is necessary to compare competing models (Sarstedt et al., 2022). Lin et al. (2017) demonstrated in a simulation study that for this purpose the information criteria *Scaled unit information prior BIC* (SPBIC) and *Haughton’s BIC* (HBIC) may be most suitable. To assess the quality of the measurement, we rely on different measures. A McDonald’s ω greater than .7 points to a sufficient internal consistency reliability; an average variance extracted (AVE) greater than .5 might indicate convergent validity, and a heterotrait–monotrait ratio smaller than 0.85 discriminant validity (Sarstedt et al., 2022). After having identified the optimal model, we extract factor scores for the subsequent latent profile analysis (Scherer et al., 2021). Factor scores are superior to sum scores (McNeish & Wolf, 2020). We opt for Bartlett factor scores as they yield unbiased estimates for the true scores (DiStefano et al., 2009). The disadvantage of factor scores is the lack of a meaningful zero point. They are standardized in a way that the sum of all students' factor scores of a construct is zero. However, this zero point does not correspond with the neutral scale mean of the five-point scale of rating (= 3). Therefore, we recalibrate the factor scores: a factor score of zero in our study means that a student is undecided concerning all the questions. Hence, we can answer whether student acceptance of a specific scenario, e.g., social robots as presenters, is positive or negative.

4.3 Latent Profile Analysis (RQ2)

The approach of the latent profile analysis follows Guggemos et al. (2022). We use the 'tidyLPA 1.1.0' R-package in combination with MPlus 8 to identify profiles of social robot acceptance by means of a latent profile analysis (LPA) (Hallquist & Wiley, 2018; Rosenberg et al., 2018). We have to restrict variances to be equal across profiles and the covariance among the variables to be zero in order to avoid convergence problems (Meyer & Morin, 2016). Missing data are not present due to the use of full information maximum likelihood. The critical step in the LPA is to identify an appropriate number of profiles. This decision might be based on information criteria and likelihood ratio tests, as well as on conceptual deliberations (Scherer et al., 2021). Against this backdrop, we first assessed different class solutions. Following Morin and Marsh (2015) and Hofmans et al. (2020), we report the information criteria AIC, CAIC, BIC, and aBIC, as well as the bootstrap likelihood ratio test (BLRT). Since these criteria may point to a different number of optimal profiles, we also utilize the approach of Akogul and Erisoglu (2017) where information criteria are weighted to determine the optimal number of latent profiles (from an empirical point of view). The identified solution should have a sufficiently high precision of classification, indicated by an entropy greater than 0.7. To demonstrate the robustness of the findings we conduct a replication of the LPA with 100 random samples of 150 students from our overall sample of 361 students (Vanslambrouck et al., 2019). Besides this, the profiles should be substantially different from each other, which can be checked by means of a MANOVA (Tondeur et al., 2019). Afterwards, we evaluate if this approach yields a meaningful solution. The profiles should be of reasonable size and show substantial shape differences, i.e., differences that not only differ in levels but also in their pattern (Morin & Marsh, 2015).

5. RESULTS

5.1 Confirmatory Factor Analysis, Model Assessment, Factor Extraction (RQ1)

The fit of the tested models is reported in Table 2. First, we tested a basic model (model 1) where the eight scenarios of robot use are operationalized with four items (see Figure 1) addressing *use intention* (use), *performance expectancy* (pe), *effort expectancy* (ee), and *ethical approval* (eth). The fit is poor: TLI = 0.858, RMSEA = 0.087. The modification indices pointed to an additional factor for ethical approval. In model 2 (see Figure 2), we added the factor *ethical approval* where, in addition to model 1, the factor ethical approval is operationalized with the ethical approval items of the eight scenarios; the fit is good: TLI = 0.954, RMSEA = 0.049. Model 3 includes, in addition to model 2, two higher order factors to consider the role of the robot, i.e., robot to class (vignettes 1–4) and robot to student (vignettes 5–8). The fit of model 3 is good: TLI = 0.954, RMSEA = 0.050. Model 4, in addition to model 2, considers the four ICAP dimensions as higher order factors: I (vignettes 4 and 8), C (vignettes 3 and 7), A (vignettes 2 and 6), and P (vignettes 1 and 5). The fit of model 4 is good: TLI = 0.953, RMSEA = 0.050. Model 5 is a combination of models 3 and 4, and considers the two roles of the robot and the four ICAP dimensions as higher order factors. The fit of model 5 is good: TLI = 0.955, RMSEA = 0.049. Based on the fit values, model 1 can be rejected. SPBIC and HBIC both point to model 2 as the best trade-off between model complexity and accuracy. Hence, we use model 2 for all our further analyses.

Table 2. Comparison of competing models

Model	Considered factors	YB–χ^2 (df)	TLI	RMSEA	SPBIC	HBIC
1	8 scenarios	1461 (436)	0.858	0.087	27,583	27,271
2	8 scenarios, 1 ethical approval	744 (420)	0.954	0.049	26,747	26,409
3	8 scenarios, 1 ethical approval, 2 1:class vs. 1:1 setting	793 (445)	0.953	0.050	26,813	26,514
4	8 scenarios, 1 ethical approval, 4 ICAP dimensions	777 (438)	0.953	0.050	26,790	26,483
5	8 scenarios, 1 ethical approval, 2 1:class vs. 1:1 setting, 4 ICAP	750 (427)	0.955	0.049	26,759	26,428

YB–χ^2 = Yuan Bentler χ^2, df = degrees of freedom, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, SPBIC = scaled unit information prior BIC, HBIC = Haughton's BIC.

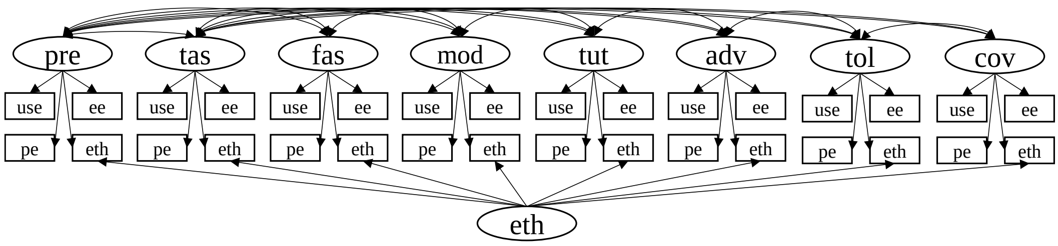


Figure 2. Preferred model (see Table 2). Note: For reasons of clarity, the correlations between the eight roles (pre – cov) with eth are not depicted. All correlations among constructs are reported in Table 3

Model 2 yields overall decent properties: YB-χ^2 (420) = 744; CFI = 0.961; TLI = 0.954; RMSEA = 0.049, 90% CI [0.041, 0.055]; SRMR = 0.029. The internal consistency reliabilities are greater than .85. The AVE is greater than .65 with one exception: the factor ethical approval shows a low AVE (.23). However, this is an auxiliary factor to consider the residual correlations among the ethical approval questions. The highest heterotrait–monotrait ratio (between *tutor* and *advisor*) equals 0.60, which is well below the cut-off value of 0.85. Table 3 also indicates moderate correlations in technology acceptance within the eight scenarios. Hence, indeed, they may capture different constructs.

As can be seen from Table 3, the overall acceptance of social robots as teaching assistants, tutors, advisors, and tools is significantly positive; the overall evaluation of robots as presenters and formative assessors is significantly negative.

Table 3. Mean, standard deviation, reliability, average variance extracted, and correlations among constructs (N = 361)

Variable	*M*	*SD*	ω	AVE	1	2	3	4	5	6	7	8
1. presenter	**-0.32**	1.08	.91	.67								
2. teaching assistant	**0.31**	1.01	.92	.67	**.41**							
3. formative assessor	**-0.12**	1.11	.92	.69	**.29**	**.47**						
4. moderator	0.01	1.10	.91	.66	**.40**	**.41**	**.25**					
5. tutor	**0.13**	1.14	.91	.71	**.24**	**.44**	**.37**	**.38**				
6. advisor	**0.26**	1.04	.91	.71	**.32**	**.44**	**.36**	**.41**	**.54**			
7. tool	**0.53**	0.90	.89	.66	**.36**	**.47**	**.30**	**.44**	**.43**	**.46**		
8. conversation partner	-0.04	1.17	.90	.71	**.33**	**.34**	**.25**	**.45**	**.38**	**.48**	**.39**	
9. ethical approval	**0.14**	0.61	.86	.23	.03	**.18**	**.12**	.03	.05	.05	.06	-.02

Note. Figures in bold indicate significance (two-sided) at the 5% level. Constructs are based on Bartlett factor scores. The factor scores are standardized such that zero equals the neutral scale mean of the five-point scale of rating.

5.2 Latent Profile Analysis (RQ2)

Table 4 shows the results of the latent profile analysis; the information criteria point to different class solutions. The algorithm of Akogul and Erisoglu (2017) suggests four profiles as the optimum. Moreover, the entropy for this solution is sufficiently high (0.813). A replication of the latent profile analysis with 100 samples of 150 students points to three profiles. In 1% of the cases, two profiles are optimal (criterion: Akogul & Erisoglu, 2017); in 57% of the cases, three profiles; in 30% of the cases, four profiles; in 12% of the cases, five or six profiles. The MANOVA yielded significant different means among the four profiles: F (24, 1015.7) = 50.16, Wilk's $\Lambda = 0.104$, $p < .001$, $\eta^2 = .400$. In other words, 40.0% of the variance can be explained by profile membership. These findings and Figure 3 point to sufficiently distinct profiles.

Table 4. Information criteria, entropies, and BLRT results for the seven latent profiles

No. profiles	-LL	No. Par.	AIC	CAIC	BIC	aBIC	Entropy	p(BLRT)
1	4,276	16	8,584	8,585	8,646	8,595	1.000	-
2	3,971	25	7,991	7,995	8,089	8,009	0.812	.000
3	3,886	34	7,841	7,848	7,973	7,865	0.790	.000
4	3,860	43	7,807	7,819	7,974	7,838	0.813	.000
5	3,841	52	7,786	7,804	7,988	7,823	0.786	.000
6	3,823	61	7,768	7,794	8,006	7,812	0.793	.000
7	3,804	70	7,749	7,783	8,021	7,799	0.780	.158

Note: -LL = - Log-likelihood, No. Par. = number of estimated parameters, AIC = Akaike's information criterion, CAIC = Consistent AIC, BIC = Bayesian information criterion, aBIC = adjusted BIC, BLRT = bootstrap likelihood ratio test.

The four latent profiles can be described as follows:

- Profile 1: Students in this profile are skeptical about social robots in general. They do not accept any of the eight social robot roles. Profile 1 comprises 45 students (12.5%).
- Profile 2: Students are skeptical about social robots as a means of instruction (scenarios 1–6 and 8). However, they assess the use of social robots as the instruction content as positive, e.g., to learn computational thinking. Profile 2 is the smallest one with 26 students (7.2%).
- Profile 3: Students have a differentiated view. They show a significant negative acceptance for social robots as presenter and formative assessor, a neutral assessment towards social robots as moderator, tutor, and conversational agent, and a significant positive assessment towards social robots as teaching assistant, advisor, and tool. Profile 3 is the largest one with 169 students (46.8%).
- Profile 4: Students have a positive acceptance towards all eight roles of social robots. This profile is also substantial in size with 121 students (33.5%).

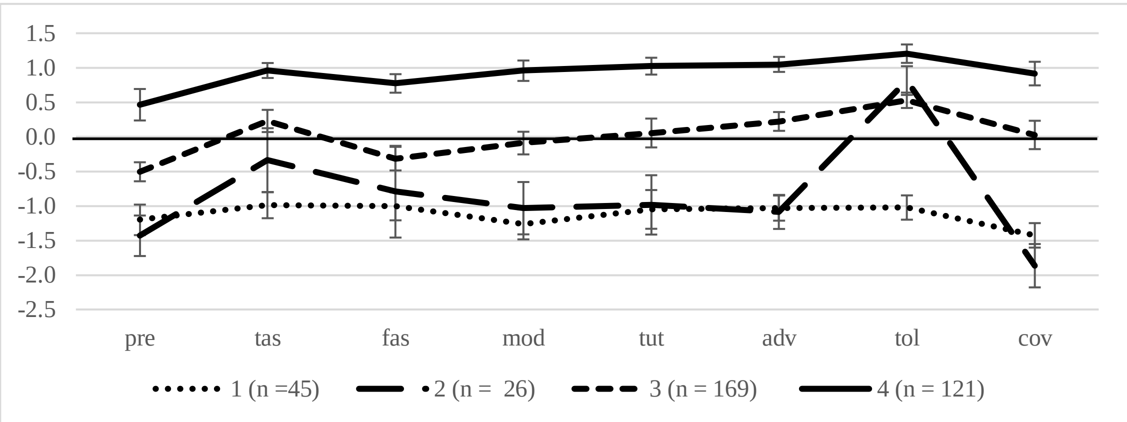


Figure 3. Latent profiles of social robot acceptance among higher education students (N = 361). Recalibrated factor scores (zero = neutral scale mean) and corresponding 95% confidence intervals

6. DISCUSSION, LIMITATIONS, AND CONCLUSION

We have developed and validated eight vignettes for social robot use in higher education. These vignettes may offer the students a fair representation of the capabilities of social robots at the current state of the art. Based on confirmatory factor analyses, there is evidence for the model structure with eight robot-use scenarios. The students regard the scenarios in substantially different and unique ways. However, when considering the four ICAP dimensions as higher order factors, this reduces the fit of the model in such a way that does not compensate the increased degrees of freedoms. The same holds true when considering the setting: robot to class vs. robot to student. From an empirical point of view, the ICAP and setting of robot use may not be a suitable structure for the vignettes. Nevertheless, they may be important from a conceptual point of view. The ICAP is a suitable framework to classify the quality of educational technology use (Sailer et al., 2021). The setting is important from a technical point of view as a robot-to-class setting is (except for the role of presenter) more difficult to implement than a robot-to-student setting.

An additional factor, *ethical approval*, is necessary in the confirmatory factor analysis in order to achieve a good model fit. The explanation for this may be due to the fact that ethical approval is a construct outside the UTAUT framework and, therefore, of a different nature. However, ethical approval may be an important facet of social robot acceptance in higher education: the AVE is greater than 0.5 for all eight scenarios; the standardized loadings of the ethical approval items on the factors exceed in all cases 0.613 ($p < .001$).

The latent profile analysis pointed to four profiles. These profiles can be interpreted in a meaningful way. However, three profiles rather than four may also be acceptable. Our replication with 100 random samples of 150 students pointed to three profiles. Moreover, the BIC—a suitable criterion to identify the optimal number of classes in mixture modeling (Nylund et al., 2007)—indicates three profiles. In the case of opting for three profiles, profiles 1 and 2 collapse. Profile 2 is small; however, it may be revealing from a conceptual point of view. Social robots can be used as a means of instruction or as the content matter in education. Students in profile 2 distinguish between these two kinds of use. We also tested models with five or more classes; however, they yield profiles that are very small and/or difficult to interpret. As a limitation, due to convergence problems, we had to assume equal variances across the profiles and covariances of zero. These assumptions are restrictive (Scherer et al., 2021). Against this backdrop, it would be beneficial to replicate our four-class solution with another sample.

Overall, the roles with the highest acceptance are social robots used as a tool and as teaching assistant. Social robots as a presenter are rated most negatively; the acceptance of social robots as a formative assessor is also negative. The vignettes may help students to obtain a more realistic picture about the capabilities of social robots in specific roles; however, they present scenarios where everything works well, i.e., no breakdowns or malfunctions of the robot occur. When social robots are used as a means of instruction (scenarios 1–6 and 8), malfunction may decrease acceptance for all roles at the same rate. Hence, the patterns of the latent classes might be unchanged but at an overall lower level. The acceptance reported in the study at hand could be regarded as the upper band. It allows conclusions about the acceptance of social robot roles that are, even in an error-free scenario, not positive. Further research could shed light on the latent profiles of students with varying degrees of acceptance toward the malfunctioning of robots.

We followed the suggestion of Skilling and Stylianides (2020) and aimed to have a vignette word count of between 50 and 200. For social robots, which are a novel and complex technology, this may not be enough to present a fair picture. On the other hand, however, more elaborate descriptions could yield higher rates of missing values and a decreasing number of participants. Overall, we regard qualitative research, which allows the study participants an in-depth interaction, as a promising way to complement quantitative studies (Sonderegger et al., 2022). In particular, the concept of ethical approval should be investigated more elaborately than has been done in our study.

Our contributions to the literature are: 1) We have developed theoretically based (ICAP framework) vignettes that fairly portray the use scenarios of social robots in higher education at the current technical state of the art. 2) We considered ethical approval as a facet of social robot acceptance, which goes beyond the classical technology acceptance variables. 3) We offer a detailed picture of student social robot acceptance by considering variance within social robot use, i.e., eight scenarios, and within students, i.e., four latent profiles.

Our work has also implications for practice. There is a considerable proportion of students (in our case, profiles 1 and 2, comprising 20% of our sample) who are skeptical about social robots. Technology acceptance should not be taken for granted among students within the social sciences. Our findings can help developers in the trade-off between technology acceptance and pedagogical and technical aspects. If the pedagogically desirable (ICAP framework) scenarios of 4 and 8 were introduced, students may be neutral towards them. It may be beneficial to accompany the introduction to such scenarios with a description about the benefits of social robot use in these roles, with a special focus on students from profiles 1 and 2.

REFERENCES

Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy*, *19*(9), 452. https://doi.org/10.3390/e19090452

Alnajjar, F., Barntneck, C., Baxter, P., Belpaeme, T., Cappuccio, M., Di Dio, Cinzia, Eyssel, F., Handke, J., Mubin, O., Obaid, M., & Reich-Stiebert, N. (2021). *Robots in education: An introduction to high-tech social agents, intelligent tutors, and curricular tools*. Routledge. https://doi.org/10.4324/9781003142706

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *Proceedings / RO-MAN 2004, 13th IEEE International Workshop on Robot and Human Interactive Communication: September 20 - 22, 2004, Kurashiki, Okayama, Japan* (pp. 591–594). IEEE Operations Center. https://doi.org/10.1109/ROMAN.2004.1374827

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), 1-9. https://doi.org/10.1126/scirobotics.aat5954

Belpaeme, T., & Tanaka, F. (2021). Social robots as educators. In OECD (Ed.), *Digital Education Outlook 2021* (pp. 143–157). OECD. https://doi.org/10.1787/7fbfff45-en

Byrne, B. M., Rossi, A., & Doolan, M. (2017). Humanoid robots attending lectures. In *Proceedings of the 9th International Conference on Education Technology and Computers, ICETC*, Barcelona, Spain.

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Ching, Y.-H., Hsu, Y.-C., & Baldwin, S. (2018). Developing computational thinking with educational technologies for young learners. *TechTrends*, *62*(6), 563–573. https://doi.org/10.1007/s11528-018-0292-7

Cooney, M., & Leister, W. (2019). Using the Engagement Profile to design an engaging robotic teaching assistant for students. *Robotics*, *8*(1), 1–26. https://doi.org/10.3390/robotics8010021

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-Determination Theory in Work Organizations: The State of a Science. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(1), 19–43. https://doi.org/10.1146/annurev-orgpsych-032516-113108

DiStefano, C., Zhu, M., & Mîndrilã, D. (2009). Understanding and using factor scores: considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, *14*(20), 1–11. https://doi.org/10.7275/DA8T-4G52

Engwall, O., Lopes, J., & Åhlund, A. (2021). Robot Interaction Styles for Conversation Practice in Second Language Learning. *International Journal of Social Robotics*, *13*(2), 251–276. https://doi.org/10.1007/s12369-020-00635-y

Fridin, M., & Belokopytov, M. (2014). Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Computers in Human Behavior*, *33*, 23–31. https://doi.org/10.1016/j.chb.2013.12.016

Guggemos, J., Seufert, S., & Román-González, M. (2022). Computational Thinking Assessment – Towards More Vivid Interpretations. *Technology, Knowledge and Learning*. Advance online publication. https://doi.org/10.1007/s10758-021-09587-2

Guggemos, J., Seufert, S., & Sonderegger, S. (2020). Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology*, *51*(5), 1864–1883. https://doi.org/10.1111/bjet.13006

Hallquist, M. N., & Wiley, J. F. (2018). Mplusautomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

Handke, J. (2018). Digitale Hochschullehre – Vom einfachen Integrationsmodell zur Künstlichen Intelligenz. In U. Dittler & C. Kreidl (Eds.), *Hochschule der Zukunft* (pp. 249–263). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20403-7_15

Handke, J. (2020). *Humanoide Roboter: Showcase, Partner und Werkzeug.* Tectum Wissenschaftsverlag.

Hofmans, J., Wille, B., & Schreurs, B. (2020). Person-centered methods in vocational research. *Journal of Vocational Behavior*, *118*, 103398. https://doi.org/10.1016/j.jvb.2020.103398

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Lin, L.-C., Huang, P.-H., & Weng, L.-J. (2017). Selecting Path Models in SEM: A Comparison of Model Selection Criteria. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(6), 855–869. https://doi.org/10.1080/10705511.2017.1363652

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Meyer, J. P., & Morin, A. J. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior*, *37*(4), 584–612. https://doi.org/10.1002/job.2085

Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 39–59. https://doi.org/10.1080/10705511.2014.919825

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., & Dong, J.-J. (2013). A Review of the applicability of robots in education. *Technology for Education and Learning*, *1*(1), 1–7. https://doi.org/10.2316/Journal.209.2013.1.209-0015

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Raffaghelli, J. E., Rodríguez, M. E., Guerrero-Roldán, A.-E., & Bañeres, D. (2022). Applying the UTAUT model to explain the students' acceptance of an early warning system in Higher Education. *Computers & Education*, *182*, 104468. https://doi.org/10.1016/j.compedu.2022.104468

Robitzsch, A. (2020). Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Frontiers in Education*, *5*, Article 589965. https://doi.org/10.3389/feduc.2020.589965

Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C., & Schmidt, J. (2018). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *3*(30), 978. https://doi.org/10.21105/joss.00978

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sailer, M., Stadler, M., Schultz-Pernice, F., Franke, U., Schöffmann, C., Paniotova, V., Husagic, L., & Fischer, F. (2021). Technology-related teaching skills and attitudes: Validation of a scenario-based self-assessment instrument for teachers. *Computers in Human Behavior*, *115*, 106625. https://doi.org/10.1016/j.chb.2020.106625

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing*, *39*(5), 1035–1064. https://doi.org/10.1002/mar.21640

Scherer, R., Howard, S. K., Tondeur, J., & Siddiq, F. (2021). Profiling teachers' readiness for online teaching and learning in higher education: Who's ready? *Computers in Human Behavior*, *118*, 106675. https://doi.org/10.1016/j.chb.2020.106675

Scherer, R., Siddiq, F., & Tondeur, J. (2020). All the same or different? Revisiting measures of teachers' technology acceptance. *Computers & Education*, *143*, 103656. https://doi.org/10.1016/j.compedu.2019.103656

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., & Paiva, A. (2017). The case of classroom robots: teachers' deliberations on the ethical tensions. *AI & SOCIETY*, *32*(4), 613–631. https://doi.org/10.1007/s00146-016-0667-2

Seufert, S., & Spiroudis, E. (2017). Wissenschaftliche Schreibkompetenz in der Studien-eingangsphase fördern: Erfahrungen mit einem Rückmelde- und Bewertungsraster (Rubric). *Zeitschrift Schreiben*, *15*, 1–9.

Sharkey, A. J. C. (2016). Should we welcome robot teachers? *Ethics and Information Technology*, *18*(4), 283–297. https://doi.org/10.1007/s10676-016-9387-z

Short, E., Sittig-Boyd, K., & Mataric, M. J. (2016). Modeling moderation for multi-party socially assistive robotics. In *IEEE Int. Symp. Robot Hum. Interact. Commun.(RO-MAN 2016).* IEEE.

Skilling, K., & Stylianides, G. J. (2020). Using vignettes in educational research: a framework for vignette construction. *International Journal of Research & Method in Education*, *43*(5), 541–556. https://doi.org/10.1080/1743727X.2019.1704243

Smakman, M. H. J., Konijn, E. A., Vogt, P., & Pankowska, P. (2021). Attitudes towards social robots in education: Enthusiast, practical, troubled, sceptic, and mindfully positive. *Robotics*, *10*(1), 24. https://doi.org/10.3390/robotics10010024

Sonderegger, S., Guggemos, J., & Seufert, S. (2022). How Social Robots Can Facilitate Teaching Quality – Findings from an Explorative Interview Study. In W. Lepuschitz, M. Merdan, G. Koppensteiner, R. Balogh, & D. Obdržálek (Eds.), *Lecture Notes in Networks and Systems. Robotics in Education* (Vol. 515, pp. 99–112). Springer International Publishing. https://doi.org/10.1007/978-3-031-12848-6_10

Tondeur, J., Scherer, R., Baran, E., Siddiq, F., Valtonen, T., & Sointu, E. (2019). Teacher educators as gatekeepers: Preparing the next generation of teachers for technology integration in education. *British Journal of Educational Technology*, *50*(3), 1189–1209. https://doi.org/10.1111/bjet.12748

Vanslambrouck, S., Zhu, C., Pynoo, B., Lombaerts, K., Tondeur, J., & Scherer, R. (2019). A latent profile analysis of adult students' online self-regulation in blended learning environments. *Computers in Human Behavior*, *99*, 126–136. https://doi.org/10.1016/j.chb.2019.05.021

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Woo, H., LeTendre, G. K., Pham-Shouse, T., & Xiong, Y. (2021). The use of social robots in classrooms: A review of field-based studies. *Educational Research Review*, *33*, 100388. https://doi.org/10.1016/j.edurev.2021.100388

Żarkowski, M. (2019). Multi-party turn-taking in repeated human–robot interactions: An interdisciplinary evaluation. *International Journal of Social Robotics*, *11*(5), 693–707. https://doi.org/10.1007/s12369-019-00603-1

Studie 7:

Seufert, S. & Guggemos, J. (2021). Neue Formen der Lernortkooperation mithilfe Künstlicher Intelligenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik* (Beiheft 31), 183–214.

Neue Formen der Lernortkooperation mithilfe Künstlicher Intelligenz

SABINE SEUFERT / JOSEF GUGGEMOS

New forms of learning location cooperation using artificial intelligence

Kurzfassung: Lernortkooperation zwischen schulischen und praktischen Ausbildungseinheiten gilt als ein zentraler Erfolgsfaktor in der Berufsbildung. Ein Forschungsdesiderat in diesem Zusammenhang ist die Veränderung der Lernortkooperation aufgrund der digitalen Transformation. Dieser Beitrag adressiert die Forschungsfrage, welche (neuen) Möglichkeiten entstehen, die Lernortkooperation in der Berufsbildung durch eine fortgeschrittene Digitalisierung, insbesondere durch Künstliche Intelligenz (KI), zu stärken. Der Beitrag legt das theoretische Fundament für veränderte Paradigmen der Lernortkooperationen. Aufbauend hierauf werden die Lernenden in den Mittelpunkt einer lernortintegrierenden Kompetenzentwicklung gestellt. Weiterhin entwirft der Beitrag Zukunftsszenarien sowie Gestaltungsoptionen, indem auf die Nutzenpotenziale von KI für Lernortkooperation eingegangen wird. Ein Reifegradmodell zeigt Entwicklungsstufen hin zu einem digitalen Ökosystem auf.
Schlagworte: Künstliche Intelligenz (KI), Lernortkooperation, Schule-Arbeitsplatz Konnektivität, Grenzobjekte, Digitales Ökosystem, Digitales Reifegradmodell

Abstract: Cooperation between learnings locations, i. e., between school-based and practical training units, is regarded as an important factor for successful vocational education and training (VET). A research desideratum in this context is the changes in learning location cooperation due to the digital transformation. The paper at hand addresses (new) opportunities to strengthen learning location cooperation in VET by means of advanced technology, in particular artificial intelligence (AI). To this end, it lays the theoretical foundation for a paradigm shift in learning location cooperation. Based on this, the learner is put in the center of the competence development across learnings locations. Furthermore, the article outlines future scenarios and design options by considering the potential benefits of AI for learning location cooperation. A maturity model shows development stages towards a digital ecosystem.

Keywords: Artificial Intelligence (AI), cooperation of locations of learning, School-Workplace-Connectivity, Boundary Objects, Digital Ecosystem, Digital Maturity Model

1 Einführung

Die Corona-Pandemie hat die berufliche Bildung hart getroffen, wie die OECD STUDIE „Bildung auf einen Blick" aufzeigt (OECD, 2020, S. 12). Praktische und betriebliche Ausbildungseinheiten konnten häufig aufgrund von Lockdown-Bestimmungen nicht durchgeführt werden. Auch gegenwärtig sind diese aufgrund von Hygiene-Vorschriften und Abstandsregelungen nur unter erschwerten Bedingungen möglich. In berufsbildenden Schulen konnten zudem häufig Qualifikationsverfahren nicht stattfinden, da Kompetenzen zur Realisierung digitaler Prüfungen fehlen. Während des Lockdowns wurde zudem deutlich, wie stark die Gesellschaft von Sektoren wie dem verarbeitenden Gewerbe und dem Gesundheitswesen abhängig ist, die sich maßgeblich auf berufliche Ausbildung stützen. Die OECD STUDIE (2020) zeigt ferner auf, dass leistungsstarke Berufsbildungssysteme ein wirksames Instrument sein können, um die Lernenden in den Arbeitsmarkt zu integrieren, die Voraussetzung für eine niedrige Jugendarbeitslosigkeit zu schaffen und somit Entwicklungsperspektiven für erfolgreiche Berufsbiographien bieten.

Im Gegensatz zu rein schulischen Ausbildungen bieten kombinierte schulische und betriebliche Bildungsgänge den Lernenden intensive Einblicke in die Arbeitswelt. Sie ermöglichen den Erwerb zukunftsbefähigender Kompetenzen, die auf dem Arbeitsmarkt nachgefragt werden (WETTSTEIN, SCHMID & GONON, 2014). In dualen Berufsbildungssystemen, wie sie in den DACH Ländern (Deutschland, Österreich und Schweiz) eine lange Tradition haben, gilt eine gelingende Lernortkooperation (LOK) als eine wesentliche Voraussetzung für eine hohe Ausbildungsqualität (WENNER, 2018; DEHNBOSTEL, 2020). Das umfasst Lernorte in verschiedenen Erfahrungsbereichen – Betrieb, Unternehmen sowie (Hoch-)Schulen (APREA, SAPPA & TENBERG, 2020). Erforderlich ist die Koordination, Integration und Zusammenarbeit aller Akteure in der Berufsbildung (Lernende, Berufsbildner*innen, Lehrpersonen etc.), was jedoch bislang nicht einfach zu realisieren war (vgl. GESSLER, 2017; RAUNER & PIENIG, 2015; EULER, 2004). Insgesamt vermitteln Forschungsbefunde zur LOK (PÄTZOLD, 1995, PÄTZOLD & WALDEN, 1995, EULER, 2004; EULER, 2015) sowie zu „School-Workplace Connectivity" (APREA, SAPPA & TENBERG, 2020) ein eher ernüchterndes Bild, inwiefern eine intensive Kooperation zwischen Schule und Betrieb in der Berufsbildungspraxis überhaupt gelingen kann. Vor dem Hintergrund der digitalen Transformation sind daher auch die Veränderungen der Gelingensbedingungen der LOK kritisch zu hinterfragen (ROHS, 2020; DEHNBOSTEL, 2020).

Bereits vor der Pandemie beherrschte der technologische Wandel viele öffentliche Debatten (SEUFERT, 2018). Es scheint mittlerweile kaum eine Branche zu geben, die nicht von der digitalen Transformation betroffen ist oder zumindest sich darauf vorbereitet, tiefgreifend davon betroffen zu sein. In der heutigen Diskussion werden allerdings die Begriffe Digitalisierung und digitale Transformation als Synonyme und inflationär verwendet (Surma & Kirschner, 2020). Die beiden Termini sind voneinander abzugrenzen und zunächst grundzulegen. Die Digitalisierung beschreibt einen Veränderungsprozess von analogen hin zu digitalen Daten. Nach Wahlster (2017) wird diese Veränderung als erste Digitalisierungswelle bezeichnet. Demgegenüber geht die digitale Transformation als zweite Digitalisierungswelle einen Schritt weiter und bezeichnet einen weitreichenden Veränderungsprozess in sämtlichen Gesellschaftsbereichen. Treiber ist hierbei insbesondere die künstliche Intelligenz (KI). Es geht nicht mehr nur um die digitale Verarbeitung der Daten, sondern neu auch um das Verstehen und Verwerten dieser Daten. Durch KI-Technologien wie Deep Learning werden unstrukturierte digitale Daten strukturiert und verwertbar gemacht. Diese maschinelle Datenanalyse und -auswertung kann Entscheidungs-, Optimierungs- oder eben auch Lernprozesse unterstützen. Da selbstlernende Systeme völlig neue und disruptive Dienste sowie Geschäftsmodelle ermöglichen, spricht Wahlster (2017) dieser zweiten Digitalisierungswelle im Sinne einer digitalen Transformation ein hohes Potenzial für Innovationen zu. Damit handelt es sich, präziser formuliert, um eine KI-Transformation, wie Makridakis (2017) betont. Unter KI soll hierbei bezugnehmend auf Bellman (1978) verstanden werden, dass technische Systeme Problemlösungs-, Entscheidungs- und Lernprozesse übernehmen können. In Bezug auf Bildungsprozesse ist Kern dieser Diskussion, dass es nicht mit einem additiven ‚Ergänzen' von Lernangeboten um soziales und mobiles Lernen getan ist. Vielmehr seien neue Geschäftsmodelle, ein Kulturwandel und veränderte Leistungsprozesse nötig (Seufert, Guggemos & Meier, 2019; Dittler, 2017; Hofhues & Schiefner-Rohs, 2017).

Dem Beitrag liegt die Begriffsdefinition dieser umfassenden digitalen Transformation zugrunde (Seufert, Guggemos & Sonderegger, 2020): 1) Organisationsentwicklung bezogen auf die gesamte Wertschöpfung in der Berufsbildung, um in Kooperation Bildungsprozesse zu organisieren. 2) Befähigung der Akteure der Berufsbildung, insbesondere Lernende, Lehrende sowie Berufsbildner*innen, diesen digitalen Wandel aktiv mitgestalten zu können. Einen Schwerpunkt im vorliegenden Beitrag nimmt folglich die KI-Transformation ein, um mittel- und langfristige Optionen für die Gestaltung der LOK zu explorieren und Entwicklungslinien zur Diskussion zu stellen.

2 Forschungsfrage und Methode

LOK hat eine lange Tradition in der Berufsbildungsforschung (Gessler 2017, S. 15). Aufgrund der aufgezeigten Veränderungen ergeben sich allerdings veränderte Rahmenbedingungen. Es besteht somit ein Forschungsdesiderat zu den Veränderungen der LOK im Zuge der digitalen Transformation (FASSHAUER, 2018). Dabei wird „die Rolle der lernortintegrativen Kompetenzentwicklung in einer vernetzten (Arbeits-) Welt immer zentraler“ (ROLL & IFENTHALER, 2020, S. 205). Die Forschungsfrage des vorliegenden Beitrages lautet folglich:

> *Welche (neuen) Möglichkeiten entstehen im Kontext der digitalen Transformation, Lernortkooperation, insbesondere durch künstliche Intelligenz, zu stärken?*

Zur Beantwortung der Forschungsfrage wird im ersten Schritt geklärt, was unter Lernort in der Berufsbildung zu verstehen ist. Darüber hinaus sollen zentrale Erkenntnisse der beiden Forschungsstränge zu LOK sowie School-Workplace-Connectivity herausgearbeitet werden, um eine Synopse zur normativen Ausrichtung der digitalen Transformation der LOK zu liefern. In einem weiteren Schritt wird ein Rahmenkonzept für Modelle einer LOK unter Nutzung von KI aufgezeigt. In diesem Zusammenhang stellt sich die Frage, wie ein Entwicklungspfad hin zu einem angestrebten Grad an LOK aussehen kann. Hierzu werden in Anlehnung an Reifegradmodelle zwei Entwicklungsstufen der digitalen Transformation abgegrenzt, die sich auf eine Literaturanalyse sowie Expertenbefragungen stützen (SEUFERT, GUGGEMOS & TARANTINI, 2018). Fluchtpunkt ist ein digitales Ökosystem im rechtlich geschützten Datenraum (SEUFERT, 2018; OSTENDORF, 2019).

Basierend auf Expertenworkshops im Projekt „FUTURE MEM“ mit dem Schweizer Verband der Schweizer Maschinen-, Elektro- und Metallindustrie (Swissmem) sowie dem vom Staatssekretariat für Bildung, Forschung und Innovation (SBFI) geförderten Projekt „Zukunftsmodelle der Lernortkooperation“ wurden erste Handlungsempfehlungen entwickelt, um die digitale Transformation der LOK mit dem Zukunftsbild eines digitalen Ökosystems für die Berufsbildung zu gestalten. In einer ersten Konzeptionsphase des geförderten Projektes wurde eine intensive Literaturstudie zu theoretischen Ansätzen und empirischen Befunden zur LOK sowie zu KI und Learning Analytics im Bildungsbereich durchgeführt. Der vorliegende Beitrag liefert Einblicke in diese umfassende Literaturstudie und den konzeptionellen Überlegungen zur Gestaltung der LOK im Rahmen der KI-Transformation.

Der vorliegende Beitrag ist folgendermaßen strukturiert:

– *Kapitel 3* dient als *theoretisches Fundament*, um die Veränderungen und das Funktionieren der Lernortkooperation in der Berufsbildung grundzulegen: 1) verändertes Verständnis von pluralistischen Lernorten und von Lernortkooperation, in der die Lernenden ins Zentrum einer integrierenden Kompetenzentwicklung gerückt werden (insbesondere relevant für personalisierte, KI-basierte Bildungsprozesse), 2) empirische Befunde zur Lernortkooperation im dualen Berufsbildungssystem, die insbesondere die institutionelle Zusammenarbeit der beteiligten Akteure (Berufsbildner*innen und Lehrkräfte) beleuchtet sowie 3) empirische Befunde zur international ausgerichteten

Forschung der School-Workplace-Connectivity, die eine integrative, lernortübergreifende Pädagogik konzipiert. Damit wird besonders die Konnektivität zwischen Arbeits- und Lernkontexten herausgestellt. Interessant an diesem Forschungsansatz ist, dass digitale Tools eine Mediatorenrolle zur Verknüpfung von Arbeits- und Lernkontexten einnehmen können. 4) In einem Zwischenfazit liefert ein mehrperspektivischer Bezugsrahmen (Makro-, Meso- und Mikroebene) den institutionellen Rahmen und integriert das konnektive Modell der Lernortkooperation. Dieser Bezugsrahmen dient als normative Orientierung und als Leitbild für die digitale Transformation der Lernortkooperation, ohne bereits spezifisch auf die Nutzenpotenziale der KI einzugehen.

– *Kapitel 4* stellt die Ergebnisse der Literaturstudie zu *KI* und insbesondere *Learning Analytics, Educational Data Mining* und *neue Mensch-Maschinen Interaktionen* vor, um die Nutzenpotenziale für eine lernortintegrierende Kompetenzentwicklung aufzuzeigen. Am Ende des Kapitels liefert ein Bezugsrahmen die Synopse für eine *KI-basierte Lernumgebung* zur wirksamen Gestaltung der lernortintegrierten Kompetenzentwicklung.
– *Kapitel 5* liefert mit dem Konzept des *Digitalen Ökosystems in der Berufsbildung* die notwendige Rahmung für eine *KI-basierte Lernortkooperation* in der Berufsbildung. Dabei werden analog zur ersten und zweiten Digitalisierungswelle (Wahlster, 2017) Entwicklungsstufen für die digitale Transformation konkretisiert.
– *Kapitel 6* liefert eine Zusammenfassung über den *mehrperspektivischen Bezugsrahmen zu einer KI-basierten Lernortkooperation* sowie Perspektiven für weiterführende Forschungsfragen.

Die Ergebnisse sollen dazu dienen, a) einen orientierenden Beitrag zur aktuellen Diskussion in der Forschung zu leisten und b) mittel- sowie langfristige Potenziale für die Entwicklung einer KI-basierten Lernortkooperation in einem digitalen Ökosystem zu identifizieren und Gestaltungsmaßnahmen auf der Makro-, Meso- und Mikroebene der Berufsbildung abzuleiten.

3 Theoretischer Hintergrund und Ergebnisse der Literaturanalyse

3.1 Zum Begriffsverständnis „Lernort“ und „Lernortkooperation“

Lernorte können, je nach Verständnis, als institutionelle oder räumliche Orte aufgefasst werden. Institutionelle Lernorte wären hierbei insbesondere der Lehrbetrieb, die Berufsschule und überbetriebliche Ausbildungsstätten. Der Lernort kann als eine im Rahmen des öffentlichen Bildungswesens anerkannte Bildungseinrichtung definiert werden, welche systematisch organisierte Lehrangebote anbietet (TIPPELT & REICH-CLAASSEN 2010, S. 11–19). Nach TIPPELT und REICH-CLAASSEN (2010, S. 11) sind Lernorte in einem weiteren Sinne alle räumlichen Einheiten, die Lernende pädagogisch stimulieren – sowohl im Kontext formal-organisierter Einrichtungen als auch im Rahmen informeller Lernprozesse.

Insbesondere geprägt durch die Rhetorik des Lebenslangen Lernens und die Bedeutung des informellen Lernens gewinnt die Pluralität der Lernorte eine neue Dimension (JUNG, 2010, S. 5). Bereits in den 1990er Jahren stellte TULLY (1994) eine Kompetenzentwicklung im Wandel und eine neue Lernkultur fest, welche die informelle Bildung durch Computer und digitale Medien bei Jugendlichen stark ins Zentrum rückt. BECK (1984, S. 258 f.) kritisierte am Lernortkonzept, der allein interessierende „Ort des Lernens" seien die Auszubildenden und daher wäre der Begriff ‚Lehrorte' geeigneter.

Nach ACHTENHAGEN, BENDORF und WEBER (2004, S. 77) können Lernorte des beruflichen Lernens als organisatorische Einheiten betrachtet werden, in denen mit oder ohne Anleitung Lernprozesse stattfinden. Das würde auch selbstorganisiertes Lernen in informellen Kontexten inkludieren.

EULER (2004, S. 13) erweitert den Lernortbegriff um pädagogisch gestaltete Einheiten in den Institutionen; so lassen sich beispielsweise Lehrwerkstatt, Übungsfirma oder neuerdings Lernfabriken 4.0 in unterschiedlichen Institutionen finden. Auch die Verlagerung spezifischer Lernphasen in Formen des online Lernens oder hybride Lernformen lassen sich mit einem solchen Zugang leichter erfassen. Der Begriff „Lernort" kann daher auf drei Ebenen verankert werden (vgl. auch EULER, 2004, S. 13): 1) Institution: Lehrbetriebe, überbetriebliche Bildungsanbieter, Berufsschule; 2) pädagogisch gestaltete Einheiten in den Institutionen, in online Lernphasen, Hybride Lern- oder Präsenzformen; 3) Person des Lernenden zur Verbindung von formal organisiertem, non-formalem (z. B. integriertes Dialog- und Trainingssystem im Geschäftsprozess der Arbeit) und informellem Lernen (z. B. Videotutorials in der Freizeit). Auch die soziale Vernetzung mit Lernenden in Praxisgemeinschaften nimmt eine hohe Bedeutung ein und erweitert das Verständnis von Lernorten. Vor dem Hintergrund der Corona-Pandemie konstatiert ROHS (2020, S. 10): „Der physische Lernort verliert einerseits an Bedeutung, andererseits werden wir uns aber auch seiner spezifischen Vorzüge bewusst. Damit verbunden ist eine bewusstere Abwägung, zu welchen Zwecken welche Lernorte genutzt und wie sie gestaltet werden." Auch DEHNBOSTEL (2020, S. 13) plädiert dafür, Lernorte neu zu definieren: „Lernorte sind örtlich und räumlich zusammenhängende Einheiten, in denen in formalen, nicht formalen und informellen Lernkontexten gelernt wird." Abb. 1 veranschaulicht dieses veränderte Verständnis.

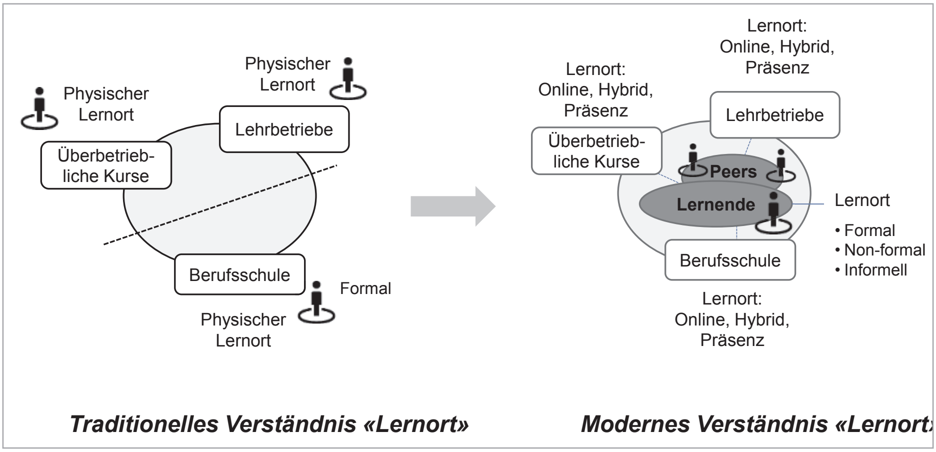


Abb. 1 Wandel des Verständnisses von Lernort und Lernortkooperation

Unter LOK wird originär ein kooperatives Bildungsmanagement verstanden, welches „das technisch-organisatorische" und „das pädagogisch begründete Zusammenwirken des Lehr- und Ausbildungspersonals, der, an der beruflichen Bildung beteiligten Lernorte", bezeichnet (Pätzold 2003, S. 72). Als normative Orientierung für die integrierende Kompetenzentwicklung im Rahmen der Lernortkooperation ist dabei der Lernende zusammen mit seinem sozialen Netzwerk ins Zentrum zu rücken, um eine integrierende Kompetenzentwicklung an unterschiedlichen Lernorten und zur Verbindung von formalem, non-formalem und informellem Lernen im Kontext des lebenslangen Lernens zu organisieren (Seufert, 2018).

3.2 Forschung zur Lernortkooperation in der dualen Berufsbildung

In der Berufsbildungsforschung hatte der vorwiegend auf das Duale System in Deutschland fokussierte Forschungsstrang zur LOK in den 1990er Jahren eine Hochphase und flachte spätestens Mitte der 2000er Jahre wieder ab (Aprea, Sappa & Tenberg, 2020). Im Modell der „dual-kooperativen Berufsbildung" wurde als Kooperationsgrundlage das Konzept der integrierten Berufsbildungspläne entwickelt (Rauner, 1998; Eder & Koschmann, 2011). Damit einher geht die Einführung von Lernfeldern bei der Strukturierung von Rahmenlehrplänen (Rauner, 1998). Die Leitidee beruflicher Bildung, anhand bedeutsamer Arbeitssituationen gleichzeitig ein Bildungsziel und die Realität der Arbeitswelt zu erfassen, bestimmt seither die Programmatik in der Berufsbildung in Deutschland, um eine neue Qualität der LOK zu etablieren (Rauner & Piening, 2015). Um die Qualitätsausprägung der LOK zu präzisieren und empirisch erfassen zu können, liegen verschiedene Typologisierungen vor:

- Typisierungsmodell nach *Intensität* (EULER, 2004, S. 14 f.) der Zusammenarbeit zwischen den Akteuren:
  1) *Informieren* der Bildungspartner: auf der niedrigsten Intensitätsstufe ist Austausch möglich sowohl im direkten Kontakt der Ausbilder*innen und Lehrkräfte als auch über die Auszubildenden oder die Berichtshefte.
  2) *Abstimmen* zwischen den Bildungspartnern als zweite Intensitätsstufe, wobei Lehrkräfte und Ausbilder*innen hier ihre Unterrichtsinhalte zwar koordinieren, jedoch unter ihren eigenen Rahmenbedingungen und Zeitplänen durchführen.
  3) *Zusammenwirken* als dritte und höchste Intensitätsstufe. Dabei werden gemeinsame Projekte durchgeführt, welche durch die Bildungspartner Berufsschule und Betrieb arbeitsteilig organisiert und umgesetzt werden.
- Typisierungsmodell nach *Kooperationsverständnis* (PÄTZOLD, 1990, S. 75 f.). Die Grundhaltungen, mit denen die Akteure einer LOK gegenüberstehen, lassen sich in vier Kategorien einordnen:
  1) *Pragmatisch-formales* Kooperationsverständnis, das ausschließlich auf äußere Einflüsse zurückgeht, wie beispielsweise die Teilnahme an Prüfungsausschüssen.
  2) *Pragmatisch-utilitaristisches* Kooperationsverständnis, als ein problemorientiertes Interesse an Kooperation, um Probleme am eigenen Lernort zu lösen.
  3) *Didaktisch-methodisches* Kooperationsverständnis, als eine Kooperation, die auf didaktischen und methodischen Überlegungen basiert.
  4) *Bildungstheoretisch-begründetes* Kooperationsverständnis, welches das didaktisch-methodische Kooperationsverständnis beinhaltet, erweitert um eine gesellschaftliche Komponente.
- Typisierungsmodell nach *Häufigkeit, Intensität und Inhalten der Kooperationsaktivitäten* (BERGER & WALDEN, 1995, S. 415–421) zwischen den Akteuren einer Lernortkooperation in fünf Stufen:
  1) *Keine* Kooperationskontakte der Bildungspartner.
  2) *Sporadische* Kooperationsaktivitäten gehen auf äußere Einflüsse zurück, beispielsweise auf die Teilnahme an Prüfungsausschüssen.
  3) *Kontinuierlich-probleminduzierte* Kooperationsaktivitäten finden nur aufgrund einzelner wahrgenommener Probleme in der Ausbildung statt.
  4) *Kontinuierlich-fortgeschrittene* Kooperationsaktivitäten finden aufgrund der Klärung von zeitlicher und/oder organisatorischer Abstimmung statt, bzw. auch, um methodisches und didaktisches Vorgehen abzustimmen.
  5) *Kontinuierlich-konstruktive* Kooperationsaktivitäten bauen auf den kontinuierlich-fortgeschrittenen Kooperationsaktivitäten auf, beinhalten aber auch Absprachen der Methoden und Inhalte. Die Kooperationsaktivitäten auf dieser Stufe können über das Informieren und Abstimmen hinaus

gehen bis hin zum Zusammenwirken in gemeinsamen Ausbildungsprojekten.

Eine Kombination der Typologisierungen liefern BERGER und WALDEN (1994), indem sie in einer Matrix das Kooperationsverständnis von PÄTZOLD (1990) den Kooperationsaktivitäten nach BERGER und WALDEN (1994) gegenüberstellen, s. Abb. 2.[1] Für die Beurteilung der praktischen Umsetzung werden drei Stufen unterschieden: wahrscheinlich, möglich und unwahrscheinlich (vgl. BERGER & WALDEN, 1994, S. 8).

Kooperationsverständnis *Kooperationsaktivität*	Pragmatisch-formal	Pragmatisch-utilitaristisch	Didaktisch-methodisch
Kooperationsabstinenz	**	0	0
Sporadisch	**	**	0
Kontinuierlich-probleminduziert	*	**	0
Kontinuierlich-fortgeschritten	0	**	*
Kontinuierlich-konstruktiv	0	**	*

0 = unwahrscheinlich * = möglich ** = wahrscheinlich

Abb. 2 Kooperationsaktivitäten und -verständnisse in der Berufsbildung (BERGER & WALDEN, 1994, S. 8)

Nach Pätzold (2003, S. 76) herrschen in der Ausbildungspraxis eher die niedrigschwelligen, pragmatischen Kooperationsverständnisse vor. Das wird durch zahlreiche empirische Studien bestätigt. Rauner und Piening (2015) untersuchten die Bewertung der Qualität der LOK durch die Auszubildenden, um zu überprüfen, inwieweit die Lernfelddidaktik eine neue Qualität der Lernortkooperation etablieren konnte. An der Befragung nahmen über 3000 Auszubildenden aus über 70 Berufen in Sachsen teil. Mit neun Items wurde die Qualität der LOK erfasst. Davon beziehen sich vier auf die Struktur der LOK, wie ‚mein Ausbildungsbetrieb und die Berufsschule stimmen die Ausbildung miteinander ab', ‚zwischen unserem Betrieb und der berufsbildenden Schule werden gemeinsame Projekte durchgeführt'. Fünf Items erheben die Qualität der Inhalte, wie ‚das Lernen in der Berufsschule und im Betrieb passt gut zusammen', ‚der Berufsschulunterricht orientiert sich an der betrieblichen Praxis', ‚die Inhalte, die ich in der Berufsschule lerne, kann ich in der Arbeit anwenden'. Die Ergebnisse der Studie zeigen, dass die Lernenden die Qualität der LOK als sehr gering einschätzen.

---

1 BERGER und WALDEN (1994) berücksichtigen den bildungstheoretischen Verständnistyp nach PÄTZOLD nicht, da dieser idealtypisch und nicht in der Praxis anzutreffen sei.

Besonders ausgeprägt zeigt sich das in kleinen und sehr kleinen Betrieben, denen die personellen Ressourcen fehlen, sich an der Gestaltung und Organisation der Zusammenarbeit mit der Berufsschule aktiv zu beteiligen (RAUNER & PIENING, 2015, S. 19). Interessant erscheint der Befund, dass die Qualität der Lernortkooperation kritischer bewertet wird, wenn die Lernenden über ein hohes Kompetenzniveau verfügen und sie die Ausbildungsqualität insgesamt (sehr) positiv bewerten.

Zu ähnlichen Ergebnissen gelangt die Studie von Wenner (2018) auf Grundlage einer Befragung von über 1.300 Auszubildenden. An den Intensitätsstufen nach EULER (2004) orientiert, beinhaltet das Erhebungsinstrument acht Items (4 Item Stufe des Informierens, 3 Items Stufe des Abstimmens, 1 Item Stufe des Zusammenwirkens). Insgesamt ist die Intensitätsstufe des Informierens am stärksten ausgeprägt, wenngleich auch diese nur ungefähr dem Skalenmittelwert entspricht. Schwächer schneidet die Stufe des Abstimmens ab. Aus Sicht der Lernenden gibt es kaum ein Zusammenwirken der beiden Lernorte. Die Aussage „Meine Berufsschule und mein Ausbildungsbetrieb führen gemeinsame Ausbildungsprojekte durch" wurde als (eher) nicht zutreffend bewertet.

Werden Lehrbetriebe befragt, ergibt sich im Vergleich zu den Auszubildenden kein abweichendes Bild. GESSLER (2017) befragte 326 Bildungsverantwortliche aus Lehrbetrieben in Bremen. Im Fragebogen waren Intensitätsstufen nach Koordination, Kooperation und Ko-Konstruktion mit insgesamt 24 Items enthalten (s. Tab. 1). Der Wert (Value) setzt sich aus zwei Faktoren zusammen: Erstens, inwieweit eine Maßnahme bereits existiert (von „1 = existiert vollständig" bis „4 = existiert nicht"), was die Notwendigkeit in der gelebten Praxis anzeigt, zweitens die Bedeutung einer Maßnahme (von „0 = unwichtig" bis „2 = wichtig"). Die Werte wurden miteinander multipliziert, um eine Rangfolge der notwendigen und gewünschten kooperativen Maßnahmen zu erhalten. Aufgrund der Bewertungen der Berufsbildner*innen ergibt sich die in Tab. 1 gezeigte Rangfolge der notwendigen und gewünschten kooperativen Maßnahmen (GESSLER, 2017, S. 189).

Tab. 1 Bedeutung von kollaborativen Aktivitäten (GESSLER, 2017)

Collaborative Measures	Value
Coordination-Level	
Exchange of information on the social behavior of apprentices	4.65
Exchange of information on the professional performance of apprentices	4.52
List of contact persons of the vocational school	4.39
Defined time slots of the availability of the contact persons of the vocational school	4.31
Exchange of information on the personal engagement of apprentices	4.25
Exchange of information on discipline and punctuality of apprentices	4.24

Collaborative Measures	**Value**
Teachers visit the training companies (exploration of the companies)	3.91
Clarification of organizational issues (e.g., examination dates)	3.23
The vocational school invites the trainers yearly to an open consultation day	3.13
Cooperation-Level	
Tuning of company training plan and school curriculum	4.12
Carry out cross-institutional learning projects	3.68
Carry out cross-institutional learn and work assignments (exploration tasks)	3.10
Joint development of training and teaching material	3.00
Conduct joint events in the vocational school	2.54
Conduct joint events in the company	2.45
Co-Construction-Level	
Internships of student teachers in the companies	3.91
Participation of company practitioners in vocational school teaching	3.33
Trainers (instructors) and teachers take part in further joint training courses	2.90
Internships of teacher in the companies	2.84
Supervising and supporting team of instructors and teachers who initiate and coordinate cooperative activities with a view to vocational school classes	2.78
Institutionalised joint working teams / joint task forces	2.73
Teachers discuss the apprentices' self reports (reports abouts the learning outcomes in the company) with the apprentices	2.52
Coordinators for cooperation at the vocational schools	2.44
Fundamental questions of cooperation are clarified in a cooperation agreement	2.28

Die Ergebnisse zeigen auch hier, dass mit zunehmender Intensitätsstufe die Werte der Aktivitäten abnehmen. GESSLER (2017, S. 189) konstatiert, dass das Potenzial der Zusammenarbeit zwischen den Akteuren – Berufsbildner*innen und Lehrkräfte – nicht ausgeschöpft ist, und sich die Zusammenarbeit in den letzten Jahren nicht wesentlich verbessert hat. Die beiden Mesosysteme Schule und Betrieb scheinen in parallelen Tätigkeitsbereichen und separater Funktionsweise ausreichend zu funktionieren.

Im Rahmen einer umfangreichen Evaluationsstudie zur Einführung der IT-Berufe untersuchten PETERSEN und WEHMEIER (2001), inwieweit sich die Einführung einer Abschlussprüfung an Berufsschulen und die Beteiligung an den Abschlussprüfungen auf die Qualität der Lernortkooperation und die Qualität der Ausbildung auswirken. Die empirischen Befunde der Studie zeigen, dass sich eine lernortübergreifende Prüfungspraxis positiv auf die Lernortkooperation auswirkt. D. h. werden also Organisationslogiken verändert, ist die Notwendigkeit zur Kooperation systembedingt und nicht vom Engagement einzelner Individuen abhängig.

Im Zuge der digitalen Transformation unterstreicht FASSHAUER (2018) die Notwendigkeit, das Potenzial der LOK stärker zu nutzen, um innovative und anspruchsvolle didaktische Lernkonzepte berufs- und lernortübergreifend zu entwickeln, wie z. B. mithilfe von Learning Analytics oder die Möglichkeiten ‚Cyber-Physischer-Systeme' (CPS) als Kooperationsgrundlage zu nutzen (vgl. hierzu näher Kapitel 4). Fundamentale Veränderungen aufgrund der fortgeschrittenen Digitalisierung bedingen, dass kaum auf elaborierte Konzepte zurückgegriffen werden kann und dass das notwendige fachliche und fachdidaktische Wissen an unterschiedlichen Lernorten vorliegt. Eine Zusammenarbeit zwischen Unternehmen und Berufsschulen erscheint deshalb dringend nötig (ebenda, vgl. hierzu ebenso ERTL, 2020).

3.3 Forschung zu School-Workplace-Connectivity

Mitte der 2000er Jahre formierte sich ein im deutschen Sprachraum bislang wenig rezipierter international ausgerichteter Forschungsstrang zur School-Workplace-Connectivity (GRIFFITHS & GUILE, 2003; TYNJÄLÄ, 2008; BILLET, 2014). Während LOK institutionell ausgerichtet und die verschiedenen Formen der Zusammenarbeit zwischen Ausbildungsakteur*innen wie Lehrkräften und betrieblichen Ausbilder*innen untersucht, setzt dieser Forschungsstrang breiter an (APREA, SAPPA & TENBERG, 2020). Der Schwerpunkt liegt dabei vermehrt auf der theoretischen Fundierung und empirischen Untersuchung von Lehr-Lern-Prozessen zur lernortintegrierenden Kompetenzentwicklung (BAARTMAN & DE BRUIJN, 2011). Der Begriff „connectivity" (Konnektivität, vgl. STENSTRÖM & TYNJÄLÄ, 2009) bezieht sich dabei auf unterschiedliche Formen der Integration von Praxisphasen und schulischen Bildungsangeboten. TYNJÄLÄ (2009, S. 11) verweist darüber hinaus auf den transformativen Charakter: „Connectivity refers to processes that contribute to close relationships and connection between different elements of learning situations, contexts of learning and systems to promote learning. Transformation refers to the changes and developmental processes to be achieved through connecting different elements of learning". Um die Verknüpfung von Lern- und Arbeitskontexten analytisch zu untersuchen, entwickelten GRIFFITHS und GUILE (2004, S. 20) fünf Modelle im Vergleich zu einem idealtypischen konnektiven Modell:

– Modell 1: Im *traditionellen Modell* besteht die Annahme, dass der Kompetenzerwerb im Betrieb nebenbei und von selbst stattfindet. Die Abstimmung der Lernorte ist minimal. Schulen stellen formale Bildungsprogramme bereit und haben keinerlei Einblick, ob ihr Bildungsauftrag von Bedeutung ist.

- Modell 2: das *erfahrungsbasierte Modell* adressiert die Notwendigkeit, dass die Lernenden spezifische Kenntnisse und Fertigkeiten sowie allgemeinere Kenntnisse und ein besseres Verständnis über den Inhalt der Arbeit erwerben müssen.
- Modell 3: im *generischen Modell* wird der Fokus auf ‚Schlüsselkompetenzen' gelegt. Der Schwerpunkt liegt auf der Nutzung von arbeitsbasierter Erfahrung zum Erwerb und zur Anerkennung von Lernergebnissen.
- Modell 4: Das *Arbeitsprozess-Modell* ist ein Versuch, die Schnittstelle zwischen der Berufsschule und dem Arbeitsplatz weiterzuentwickeln. Dabei soll ein ganzheitliches Bild der Arbeitsprozesse und -inhalte erzeugt werden. Ziel ist die Verbindung von fachtheoretischem und fachpraktischem Wissen. Die Schule fördert den Wissenstransfer, z. B. durch handlungsorientierte Unterrichtsmethoden.
- Modell 5: das *konnektive Modell* geht über die Prinzipien dualistischer Modelle hinaus und vertritt eine grundlegend andere Annahme über Lernen und Entwicklung. Danach sind alle Formen des Lernens ‚situiert' und Wissen wird in Interaktion zu einer sozialen Umwelt konstruiert und transformiert (Griffiths & Guile, 2003; Tynjälä, 2009). Der Transfergedanke im Sinne eines Anwendens von Theoriewissen in der Praxis („Paketmodell": Hautz & Ostendorf, 2020, S. 116) wird abgelehnt. Vielmehr geht es um eine „Rekontextualisierung von Wissen aus beiden Tätigkeitssystemen, um die Betrachtung von Wissen als kritisch zu hinterfragendem Werkzeug, um die beidseitige Gestaltung von Übergängen" (ebenda).

Einen weiteren Schwerpunkt im Forschungsstrang School-Workplace-Connectivity bilden die wissenschaftlichen Ansätze zu ‚Boundary Crossing' (Star & Griesemer, 1989; Akkerman & Bakker, 2011). Ziel ist es hierbei, eine integrative Pädagogik an verschiedenen Lernorten zu unterstützen. Mit dem Aufbau von sogenannten ‚Boundary Crossing Skills' ermutigen Lehrkräfte und Berufsbildner*innen die Lernenden, theoretisches Wissen, Erfahrung und Selbstregulierung beim Lernen zu kombinieren (Aprea & Sappa, 2020). Im Kontext dieser Überlegungen bieten digitale Technologien wie Apps, Blogs und Videos das Potenzial, den Lernenden dabei zu helfen, Lernen in der Schule sowie am Arbeitsplatz zu verbinden und somit mittels ‚Boundary Objects' (Caruso, Cattaneo & Gurtner, 2020) zur Förderung von Konnektivität und integrativer Kompetenzentwicklung beizutragen (Kilkbrink, Enochsson & Söderlind, 2020; Aprea et al., 2012).

Nach Star und Griesemer (1989, S. 393) sind diese ‚Boundary Objects' (Grenzobjekte) diejenigen Objekte, die mehrere sich überschneidende Welten betreffen und den Informationsbedürfnissen jeder dieser Welten entsprechen. Typisch für Grenzobjekte ist, dass sie in der gemeinsamen Nutzung schwach strukturiert und in der individuellen Nutzung der Standorte stark strukturiert sind (Akkerman & Bakker, 2011,

S. 141). Grenzobjekte können als „ein Mittel zur Übersetzung" (STAR & GRIESEMER, 1989, S. 393) innerhalb einer Situation von lernortübergreifenden Aktivitätsbeziehungen und Anforderungen betrachtet werden. Grenzobjekte können daher auch als Mediatoren betrachtet werden (ebenda). So können beispielsweise digitale Lern- und Leistungsdokumentation darin unterstützen, an unterschiedlichen Lernorten erworbene Wissensbestände zu verknüpfen (CATTANEO & APREA, 2018; APREA, CATTANEO & SAPPA, 2015).

Weitere interessante Perspektiven zur Stärkung der Konnektivität zwischen den Lernorten können sich mit Mediatoren in Form von smarten Lernräumen (z. B. Augmented Reality, Virtual Reality oder Mixed Realities) sowie auch in Form von smart machines (z. B. virtuelle Assistenten, intelligente chatbots oder soziale Roboter) ergeben (SEUFERT, GUGGEMOS & SONDEREGGER, 2020). Auf diese KI-basierten Entwicklungen wird im Kapitel 4 näher eingegangen.

Das Modell der *integrativen Pädagogik* der Forschergruppe um TYNJÄLÄ (2008; TYNJÄLÄ et al., 2016) knüpft daran an und verbindet die Kombination von theoretischem Wissen und praktischem Erfahrungslernen mittels Mediationstools und -prozessen. Kontextualisierung, Mediation, Partizipation in Praxisgemeinschaften und Konstruktion von Wissen werden dabei als wichtige Aspekte des beruflichen Kompe-

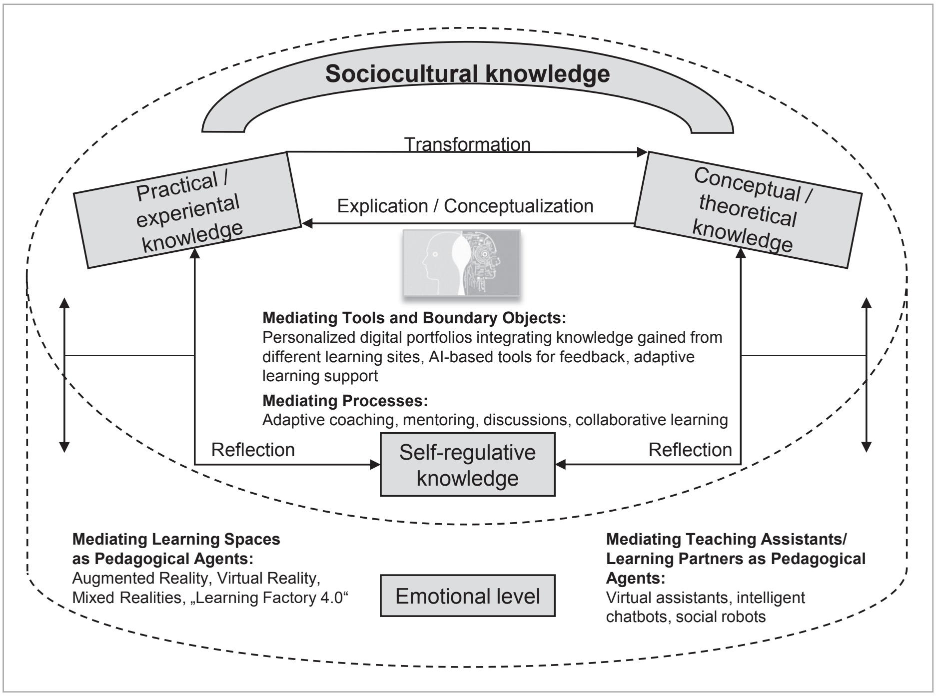


Abb. 3 Das Modell der integrativen Pädagogik (in Anlehnung an TYNJÄLÄ, 2008; TYNJÄLÄ, et al., 2016).

tenzerwerbs betrachtet (Ostendorf, 2020, S. 116). Da heute sowohl in der Schule als auch im Lehrbetrieb konzeptionelles, theoretisches Wissen erworben wird (vgl. hierzu Wettstein, Schmid & Gonon, 2014), eignet sich dieses Modell der integrativen Pädagogik in besonderem Maße für eine lernortintegrierende Kompetenzentwicklung. Abb. 3 verdeutlicht die Zusammenhänge im Modell der integrativen Pädagogik.

Zusammenfassend lässt sich festhalten, dass die Konnektivität zwischen den Lernorten eine enge Kooperationsgrundlage benötigt (Roll & Ifenthaler, 2020). Lernortübergreifende Lernräume (z. B. Lernfabriken 4.0) sowie KI-basierte pädagogische Agenten bieten neues Potenzial, um eine Basis für die intensive Zusammenarbeit zwischen den Lernorten zu liefern sowie die Konnektivität zwischen Arbeits- und Lernkontexten zu erhöhen.

3.4 Zwischenfazit: Mehrperspektivischer Bezugsrahmen und konnektives Modell zur Verknüpfung von Arbeits- und Lernkontexten

Beide Forschungsstränge, zu LOK in der dualen Berufsbildung sowie zur School-Workplace-Connectivity, betonen die Bedeutung, einen mehrperspektivischen Bezugsrahmen für die Analyse und Gestaltung der LOK heranzuziehen:

- Auf der *Makroebene bzw. Systemebene* bezieht sich diese Verknüpfung auf die Passung zwischen Bildungsinstitutionen und Arbeitswelt; diese Ebene fokussiert vor allem auf die Gestaltung der Arbeitsmarkt- und Berufsbildungspolitik. Die Konnektivität bezieht sich auf eine systemische Einbettung in das Bildungssystem und den Arbeitsmarkt.
- Die *institutionelle Ebene (Mesoebene)* betrifft die Organisation und Koordination der Interaktion und Kommunikation zwischen den Akteuren in Schule bzw. Hochschule und Betrieb. Zudem werden auf dieser Ebene die curricularen Regelwerke in den Blick genommen. Die Konnektivität zwischen den beiden Tätigkeitssystemen bezieht sich dabei auf die organisationalen Strukturen und Bedingungen von Lernumgebungen in Schule und Betrieb, um eine integrierende Kompetenzentwicklung zu gestalten.
- Auf der *instruktionalen Ebene (Mikroebene)* wird der Schwerpunkt auf die Verknüpfung von (hoch-)schulischen und betrieblichen Lehr-Lern-Prozessen gelegt. Hier geht es vorrangig darum, wie (hoch-)schulische und betriebliche Lernarrangements gestaltet werden müssen, um die integrative Kompetenzentwicklung der Lernenden (und auch des Berufsbildungspersonals) optimal zu fördern. Die Konnektivität bezieht sich dann auf den individuellen Lehr- und Lernprozess.

Inwieweit Schwierigkeiten bei der Lernortkooperation von unterschiedlichen Systemen auf der Makroebene abhängen, wird in der Literatur diskutiert. Rauner (2017) sieht in der Trennung auf der Makroebene in Deutschland (Bund zuständig für berufliche Bildung, Länder zuständig für schulische Bildung) einen maßgeblichen Grund fehlender Grundlagen für eine intensive Lernortkooperation auf der Meso- und Mikroebene. Im Vergleich dazu wird das duale Berufsbildungssystem in der Schweiz auf der Makroebene zentral gesteuert und nimmt daher nach Ansicht von RAUNER (2017) eine Vorbildfunktion ein. Es gibt allerdings auch Studien, die Schwächen der LOK, konkret der Zusammenarbeit zwischen Betrieben und Schulen, auch dem Schweizer Berufsbildungssystem attestieren (SAPPA & APREA, 2014).

Im Zuge der digitalen Transformation sind daher notwendige Kooperationsgrundlagen auch auf der Makroebene zu überdenken, um die eher separat agierenden Mesosysteme Lehrbetrieb und Schule mit einer Organisationslogik für eine integrierende Kompetenzentwicklung zu unterstützen. Für ein gemeinsames Kooperationsverständnis liefert das konnektive Modell zur Verknüpfung von Arbeits- und Lernkontexten eine geeignete Basis für Transformationsprozesse. Die Rollen der Lernorte ändern sich dabei, da transformationales Lernen sowohl auf individueller als auch auf organisationaler Ebene stattfindet. In der beruflichen Bildung rücken veränderte Praktiken in interprofessionellen Arbeits- und Lernprozessen in den Vordergrund. Der Lernort Schule erhält dabei eine neue und starke Positionierung, um Partnerschaften mit

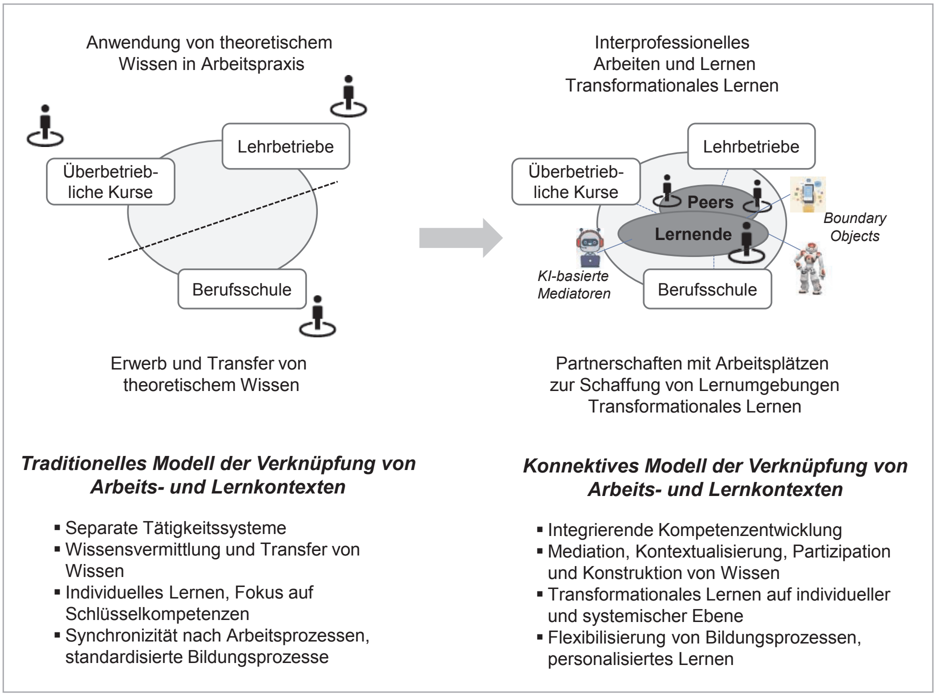


Abb. 4 Traditionelles vs. konnektives Modell der Verknüpfung von Arbeits- und Lernkontexten

Arbeitsplätzen zur Schaffung von Lernumgebungen einzugehen. Abb. 4 zeigt das konnektive Modell in Abgrenzung zum traditionellen Modell.

Dieses konnektive Modell dient als normative Orientierung für die Gestaltung der Lernortkooperation und insbesondere für die lernortintegrierende Kompetenzentwicklung. Im folgenden Abschnitt wird auf die Nutzenpotenziale der KI eingegangen, um dieses konnektive Modell zur Verknüpfung von Arbeits- und Lernkontexten in der Berufsbildung auszudifferenzieren.

4 Nutzenpotenziale der künstlichen Intelligenz für die lernortintegrierende Kompetenzentwicklung

4.1 Überblick: Paradigmenwechsel in der Organisationslogik

Bereits in den 1990er Jahren ist die Notwendigkeit eines Paradigmenwechsels in der Berufsbildung intensiv diskutiert worden (vgl. hierzu Arnold, 2020 sowie Erpenbeck & von Rosenstiel, 2007). Brater (1992, S. 85) konstatierte, dass die Berufsausbildung einen Paradigmenwechsel benötigt: „Während sie sich nämlich bisher vor allem an dem orientieren konnte, was ‚Bedarf‘ des Beschäftigungssystems war, und ihre Aufgabe hauptsächlich darin bestand, Wege zu finden, wie der einzelne an diesen Bedarf anzupassen war, so muss sich nun ihr Blick primär auf die Person des Lernenden und ihre je spezifischen individuellen Entwicklungsmöglichkeiten richten.“ Berufliche Bildung vermittelt nicht nur Kenntnisse und Fähigkeiten, sondern stiftet auch Haltungen und fördert Offenheit, Flexibilität und Wirksamkeit im gestaltenden Umgang mit neuen Anforderungen. Neben der fachlichen Kompetenzentwicklung steht zunehmend die Stärkung der ‚Persönlichkeiten‘ der Auszubildenden im Vordergrund. Komplementäre Kompetenzen zu (intelligenten) Maschinen und digitalen Systemen wie etwa Kreativität, kritisches Denken, Erfindergeist oder Empathie werden an Bedeutung gewinnen (Seufert, 2018).

Das Verfolgen dieser strategischen Leitlinien für die Berufsbildung erfordert einen Paradigmenwechsel in der Organisationslogik. Das umfasst die Planung, Durchführung, Evaluation und Steuerung von Bildungsprozessen. In diesem Kapitel wird demnach der Frage nachgegangen, wie die Entwicklungen im Bereich KI für diesen Paradigmenwechsel genutzt werden können. Drei Entwicklungslinien sollen dabei im Vordergrund stehen: 1) Learning Analytics zur wirksamen Gestaltung der lernortintegrierenden Kompetenzentwicklung, 2) Educational Data Mining zur Gestaltung von (teil-)automatisierten Bildungsprozessen sowie 3) neue Formen zur Förderung der Konnektivität durch KI-basierte Arbeits- und Lernumgebungen sowie KI-basierte Mediatoren.

4.2 Learning Analytics zur Unterstützung der Lernortkooperation

Als *Learning Analytics* wird die Interpretation verschiedenartiger Daten bezeichnet, die von Lernenden produziert oder für sie erhoben werden, um Lernfortschritte zu messen, zukünftige Leistungen zu prognostizieren und potenzielle Problembereiche aufzudecken (FERGUSON et al., 2014; IFENTHALER, 2017; SEUFERT et al., 2019). Im Vordergrund steht hierbei, die Logik von Data Science auf die Organisation von Bildungsprozessen zu übertragen (DILLENBOURG, 2017). Mithilfe der Auswertung der Daten sollen die Lernenden wirksamer und damit auch besser in ihren Lernprozessen unterstützt und dadurch der Erfolg der Berufsbildungsprozesse insgesamt gesteigert werden (IFENTHALER, 2017; BUCKINGHAM SHUM & DEAKIN CRICK, 2012). Durch individuelle, zeitnahe, präzise sowie auch kompakte Feedbacks für alle Beteiligten der Lernortkooperation ergeben sich neue Chancen für eine wesentliche Qualitätsverbesserung der Bildungsprozesse im Rahmen der Lernortkooperation (SEUFERT, 2018). Dabei ist zu betonen, dass die Interpretation der lernerspezifischen Daten bzw. die notwendige Intervention i. d. R. nicht automatisiert abläuft, sondern auf die Erfahrung von Lehrenden (Lehrpersonen, Berufsbildner*innen) zurückgreift.

GRELLER und DRACHSLER (2012, S. 46) zeigen die Informationsflüsse auf verschiedenen Systemebenen auf. Damit verbunden sind Handlungsoptionen von Akteuren auf verschiedenen Ebenen. Auf der Mikro-Ebene sind das in erster Linie die Lernenden, Lehrpersonen sowie Berufsbildner*innen. Diese können auf Grundlage solcher Informationen ihr lehr-/lernbezogenes Handeln überprüfen, reflektieren und gegebenenfalls anpassen (DILLENBOURG, 2017). Andererseits ergeben sich damit auch grundsätzlich neue Möglichkeiten der empirischen Bildungsforschung. Eine offene Frage stellt sich, welche Auswertungen lernortübergreifend nutzenbringend sein könnten und so die Aktivitäten von, beziehungsweise die Herausforderungen für verschiedene Institutionen in der Berufsbildung besser in den Blick zu nehmen (Learning Analytics auf der Meso- sowie Makro-Ebene der Berufsbildung).

4.3 Educational Data Mining zur Gestaltung von (teil-)automatisierten Bildungsprozessen

Mit *Educational Data Mining* können auf der Basis von Machine Learning Algorithmen Lernsysteme entwickelt werden, die künftig immer mehr Aufgaben in Bildungsprozessen übernehmen können (ALDOWAH, AL-SAMARRAIE & FAUZY, 2019). Damit kann einerseits eine Entlastung bei den Lehrpersonen geschaffen (z. B. weniger Korrekturarbeiten) sowie auch eine Qualitätsverbesserung von Lehr-Lernprozessen (z. B. automatisiertes Feedback bei Aufgaben auf einer höheren Lernziel-Taxonomiestufe) erzielt werden (SEUFERT, GUGGEMOS & SONDEREGGER, 2020). Um derartige KI-basierte Lernsysteme zu entwickeln, sind große Datenmengen für eine kontinuierliche

Weiterentwicklung der Systeme notwendig (Ferguson, 2014; Ifenthaler, 2017; Renz & Hilbig, 2020). Für die Berufsbildung können hier beispielsweise KI-basierte Feedbacksysteme, adaptive Lernsysteme oder KI-basierte Empfehlungen zu Lerninhalten/-wegen interessante Einsatzmöglichkeiten bieten. Wie die aktuelle Metastudie von Zawacki-Richter et al. (2019) allerdings aufzeigt, steht die Forschung zu KI im Bildungsbereich erst am Anfang; es liegen erst sehr wenig Forschungsbefunde vor.

4.4 Neue Formen der Konnektivität durch KI-Systeme

Mit der zunehmenden Verbreitung von Augment Reality (AR) und Virtual Reality (VR), d. h. immersiven Arbeits- und Lernumgebungen, werden neue, flexibel einsetzbare Formen für Trainingsprogramme direkt am Arbeitsplatz ermöglicht (Seufert, 2018). Auch in berufsbildenden Schulen sind bereits Projekte entstanden, welche die Potenziale dieser Lernumgebungen in konkreten didaktischen Szenarien explorieren und damit auch neue Formen der LOK erproben. Wenn AR- oder VR-Technologien intelligenzbasierte Elemente einbinden, so können personalisierte Rückmeldungen und Empfehlungen in einer Simulationsumgebung einen didaktischen Mehrwert für ein intensives Coaching der Lernenden bieten.

Im Kontext von Industrie 4.0 begründen CPS die technische Basis für Industrie 4.0, welche auch eine Kooperationsgrundlage zwischen den Lernorten bieten können. Als integrierte und vernetzte Systeme können sie über entsprechende Benutzerschnittstellen auch mit Menschen interagieren, so dass selbstorganisierte, effiziente und flexible Produktionsprozesse intelligent umgesetzt werden können (Scheid, 2018). In diesem Zusammenhang könnten Lernfabriken 4.0 als modellhafte Industrie 4.0 Produktionsanlagen und didaktische CPS fungieren. Einerseits kann dadurch die Kooperation zwischen gewerblichen Schulen und Betrieben, andererseits aber auch zwischen gewerblichen und kaufmännischen Schulen gestärkt werden, um berufsfeldübergreifende Kompetenzen zu fördern (Scheid, 2018).

Darüber hinaus können neue Formen der Mensch-Maschine Interaktion die Konnektivität zwischen den Lernorten fördern. Virtuelle Assistenten, intelligente Chatbots oder soziale Roboter (Guggemos, Seufert & Sonderegger, 2020) können als KI-basierte Mediatoren zwischen Arbeits- und Lernwelten fungieren. Als unermüdliche Trainingspartner können sich Lernende beispielsweise direkt mit den Lerninhalten am Arbeitsplatz ‚unterhalten'. Chatbots sind in vielen Bereichen wie der Finanzbranche und auch im Bildungswesen eine zunehmend eingesetzte Technologie, mit der viele Organisationen den Einstieg in KI-Entwicklungen wagen (Seufert et al., 2020). Dabei wird die Konversation mit menschlichen Nutzer*innen, insbesondere über das Internet, in Form von auditiven oder textuellen Gesprächen simuliert (Winkler & Söllner, 2018). KI-basierte Mediatoren bieten daher neues Potenzial für eine gemeinsame Entwicklung und Zusammenarbeit über die Lernorte hinweg und somit auch für eine lernortintegrierende Kompetenzentwicklung (RENZ, KRISHNARAJA & GRONAU, 2020).

4.5 Synopse: Bezugsrahmen einer KI-basierten Lernumgebung für eine lernortintegrierende Kompetenzentwicklung

Für eine lernortintegrierende Kompetenzentwicklung in der Berufsbildung können KI-basierte Lernumgebungen, wie beispielsweise ‚Learning Experience Plattformen', dabei unterstützen, stärker den Lernenden von der Organisationslogik her in den Mittelpunkt zu stellen. International wird als eine der bedeutendsten Vorteile des Einsatzes der KI im Bildungsbereich die Unterstützung personalisierten Lernens gesehen (STANFORD UNIVERSITY, 2016). Für die Berufsbildung ist aufgrund der systemischen Einbettung in den Arbeitsmarkt ein zunehmender Druck zu beobachten, sich intensiv mit den Entwicklungen und Implikationen der digitalen Transformation zu beschäftigen. ROHS (2020, S. 9) macht in diesem Zusammenhang auf das Spannungsfeld zwischen Personal- und Persönlichkeitsentwicklung aufmerksam: „Die Herausforderung besteht dabei darin, nicht nur die notwendigen Kompetenzen zur Ausführung einer gerade notwendigen Tätigkeit im Blick zu haben, sondern Überblickwissen zu generieren, Zusammenhänge herzustellen, die Verbindung zu den theoretischen Grundlagen deutlich zu machen und Reflexionsprozesse in Gang zu setzen. Darüber hinaus geht es weiterhin darum, Freiheiten für die individuelle Entwicklung zu schaffen, d. h. nicht nur an den Anforderungen im Arbeitsprozess anzusetzen, sondern sich auch gezielt davon zu entfernen." Aus didaktischer Sicht zieht ROHS (2020, S. 9) weiters den Schluss, dass es folglich der „selbstgesteuerten Konfiguration der Lernorte" bedarf. Das bedingt Veränderungen in den Prozessen, Strukturen und Kulturen, um mittels KI eine stärkere Personalisierung der Berufsbildungsprozesse zu erreichen.

Entlang der didaktischen Wertschöpfungskette von der Ausbildungsplanung, über Maßnahmen zum Erwerb von Handlungskompetenzen bis hin zum abschließenden Qualifikationsverfahren können personalisierte Bildungsprozesse bezogen auf Ziele, Inhalte, Methoden sowie Lernbegleitung berücksichtigt werden. Handlungsleitend sind dabei die curricularen Vorgaben zur Erzielung eines Kompetenzprofiles, das in einem Baukastensystem nach Basis- und Wahlqualifikationen zur Vertiefung sowie berufs(feld-) und berufsspezifischen Handlungskompetenzen unterscheiden kann. Eine derartige Lernumgebung würde alle relevanten Inhalte und Informationen von den verschiedenen Lernorten über eine entsprechende Schnittstelle erhalten.

Personalisiertes Lernen orientiert sich dann konsequent an den Bedürfnissen der Lernenden in Abstimmung mit den Berufsbildner*innen im Ausbildungsbetrieb sowie den Lehrpersonen. Sie definieren selbstorganisiert ihre Lernziele aufgrund aktueller Herausforderungen, planen ihre Lernprozesse eigenverantwortlich und optimieren sie laufend auf Basis der Ergebnisse und Rückmeldungen im Arbeitsprozess und zunehmend auch auf der Basis von KI-basierten Empfehlungen (z. B. personalisierte Lernwege in adaptiven Lernsystemen, Vernetzungsvorschläge mit Peers, Prognosen bzgl. Lernzeit, -tempo, Angaben zum Kompetenzfortschritt etc.). Die Lehrkräfte werden dabei nicht überflüssig, sie begleiten nunmehr die individuellen Lernprozesse, beraten im Hinblick auf Lernstrategien sowie Lerntechniken und unterstützen bei Bedarf unter Rückgriff auf KI-basierte Instrumente. KI-basierte Arbeits- und Lernumgebungen sowie Mediatoren bieten das Potenzial, eine Kooperationsgrundlage für eine lernortintegrierende Kompetenzentwicklung zur Verfügung zu stellen. Auch die selbstgesteuerte Konfiguration der Lernorte wird darüber hinaus im Bezugsrahmen einer KI-basierten Lernumgebung unterstützt. In Abb 5. sind die Zusammenhänge abschließend visualisiert.

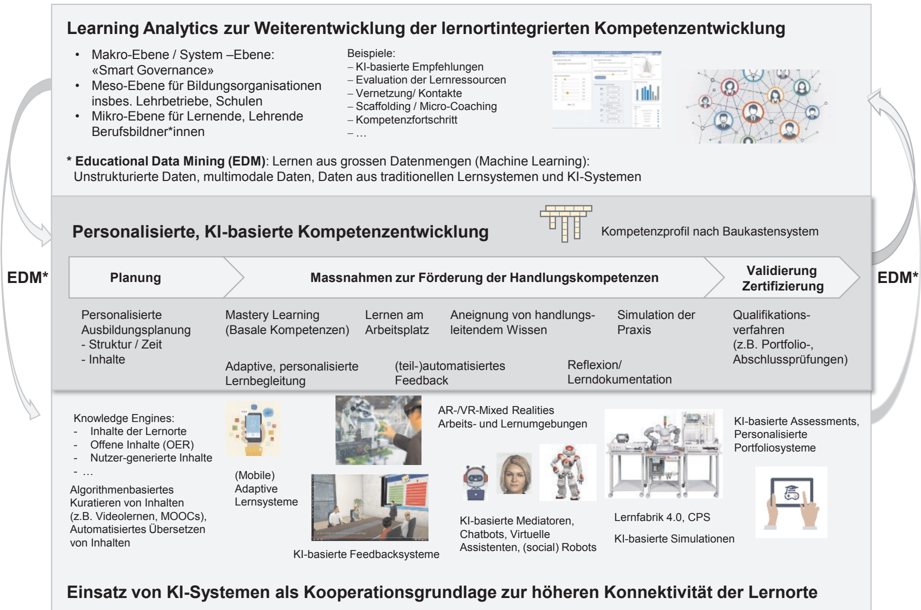


Abb. 5 Konzept einer KI-basierten Lernumgebung für eine lernortintegrierte Kompetenzentwicklung in der Berufsbildung

5 Digitales Ökosystem für neue Modelle der Lernortkooperation

Fluchtpunkt der Überlegungen für neue Modelle der LOK stellt der Aufbau digitaler Ökosysteme für die Berufsbildung dar. Digitale Ökosysteme bieten in Form von offenen Lernsystemen einen neuen Gestaltungsrahmen, um die Chancen der KI in einer Netzwerkökonomie nutzen zu können (SEUFERT, GUGGEMOS & MOSER, 2019). Für die Entwicklung und kontinuierliche Weiterentwicklung von KI-Systemen sind große Datenmengen erforderlich, die auf der Basis von Plattformökonomien gewonnen werden können. Offene Lernökosysteme können in diesem Zusammenhang auch als Bindeglied zu Open Education im digitalen Bildungsraum aufgefasst werden. KERRES und HEINEN (2015) stellen diesen Zusammenhang explizit mit dem Konzept des „Informational ecosystem" her.

In Anlehnung an Reifegradmodelle können drei Entwicklungsstufen der digitalen Transformation gebildet werden, die sich auf eine Literaturanalyse sowie Expertenbefragungen stützen (SEUFERT, GUGGEMOS & TARANTINI, 2018). Damit kann ein Entwicklungspfad hin zu einem normativ erwünschten Grad an Lernortkooperation beschrieben werden. Bei der digitalen Transformation der Berufsbildung geht es darum, diese Entwicklungsstufen zu verstehen und für die Qualitätsentwicklung der Berufsbildung zu nutzen. Mithilfe von Reifegradmodellen können die Entwicklungen fassbar und dadurch besser plan- und steuerbar gemacht werden, während gleichzeitig die langfristigen Entwicklungsziele im Blick behalten werden (SCHALLMO et al., 2017). Bisherige Reifegradmodelle beziehen sich auf die digitale Transformation einer Organisation (vgl. hierzu die systematischen Reviews von IFENTHALER & EGLOFFSTEIN, 2020 sowie THORDSEN, MURAWSKI & BICK, 2020). Die digitale Transformation der Lernortkooperation als Zusammenspiel von Netzwerkpartnern zur gemeinsamen Er-

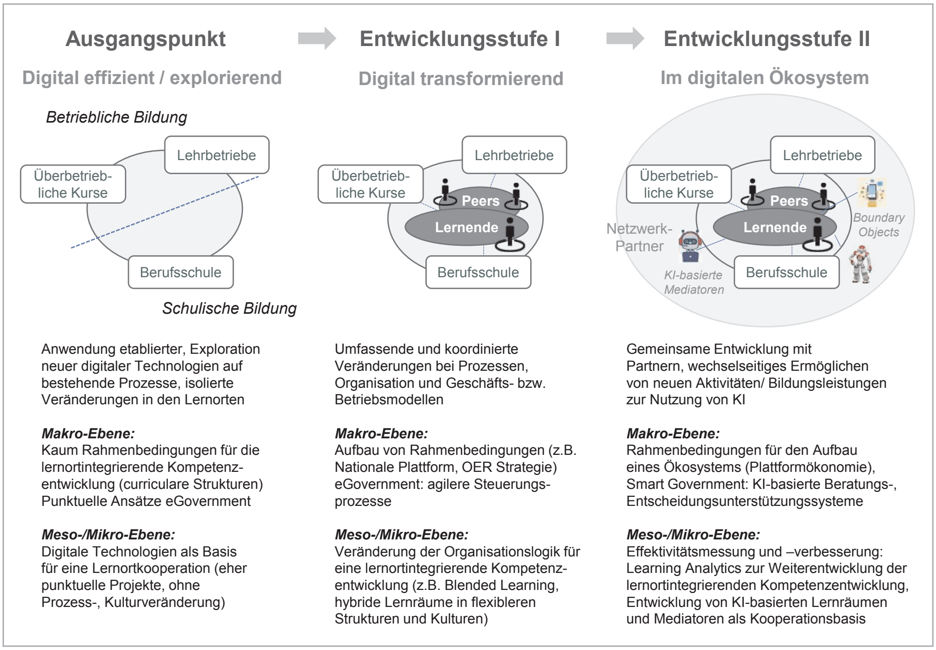


Abb. 6 Reifegrademodell der KI-Transformation der Lernortkooperation (in Anlehnung an Seufert, 2018).

bringung und kontinuierlichen Weiterentwicklung von Bildungsdienstleistungen ist bislang noch nicht erforscht.

Die Entwicklungsstufen in Abb. 6 basieren auf der empirischen Arbeit zuhanden des Staatssekretariats für Bildung, Forschung und Innovation [SBFI] (Seufert, 2018) und der differenzierten anschließenden Deliberation (Seufert, Guggemos & Tarantini, 2018; Seufert, Guggemos & Sonderegger, 2020). Durch diese Entwicklungsstufen wird der Prozess hin zu einem Ökosystem aufgezeigt, das ein personalisiertes, KI-unterstütztes Lernen für eine integrierende Kompetenzentwicklung in verschiedenen Lernorten organisiert.

Ausgangspunkt: Digital effizient/explorierend

Derzeit findet die betriebliche Bildung in Lehrbetrieben und überbetrieblichen Kursorten einerseits und die schulische Bildung andererseits jeweils in eigenen Organisationslogiken statt, was zu eher starren Rahmenbedingungen führt. Die Koordination und Synchronisierung der Ausbildungsinhalte sind schwer umsetzbar (Aprea & Sappa, 2015). So befassen sich beispielsweise die Lernenden mit Inhalten in der Schule, die erst sehr viel später relevant in ihrem Lehrbetrieb werden. Auch sind die Grundverständnisse in den Lernorten Schule und Betrieb häufig sehr unterschiedlich. Hinsichtlich der Organisationslogik bewegen sich Schulen meist in einem Kohortensystem im Klassenverbund, was den Lehrkräften das individualisierte Unterrichten in zunehmend heterogenen Klassen erschwert. Starre Rahmenbedingungen in Schulen erschweren zudem häufig den Einsatz digitaler Medien im Unterricht (z. B. fragmentierter Unterricht im 45-Minuten-Takt). Regelmäßig werden auch verschiedene Lernplattformen in den unterschiedlichen Lernorten eingesetzt. Darüber hinaus bestehen häufig große Unsicherheiten und Unklarheiten in Bezug auf rechtliche Rahmenbedingungen (z. B. Datenschutz und Persönlichkeitsrechte).

Entwicklungsstufe I: Digital transformierend

Zunächst besteht die Herausforderung, die LOK flexibler zu gestalten. Ertl (2020, S. 3) betont, dass aufgrund der derzeitigen Erfahrungen mit der Pandemie didaktische Konzepte für das Lernen in hybriden Lernräumen zu entwickeln seien: „Im Mittelpunkt muss dabei die Befähigung der Lernenden und des Bildungspersonals stehen, Lernumgebungen sinnvoll zu gestalten und an die eigenen Bedürfnisse anzupassen. Hierfür bietet das Lernen am Arbeitsplatz vielfältige Anknüpfungspunkte. Denn die Gestaltung lernförderlicher Arbeitsprozesse verbindet berufliches Handeln und Kompetenzentwicklung, wie es auch für die lernortübergreifende Entwicklung von Lernprozessen notwendig ist."

Auf Makroebene fokussiert eGovernment zunächst vor allem auf die Sicherstellung eines rechtlich geschützten Datenraums (SEUFERT, 2018). Sie trägt zur Lösung der Herausforderungen bei, denen sich das Bildungssystem im Bereich von Datenschutz und -sicherheit, Infrastruktur sowie digitalen Inhalten zu stellen hat. Der Zugriff auf individuelle Daten des Einzelnen kann damit transparent gestaltet werden und die Entscheidungsgewalt über die Nutzung der Daten bleibt bei den Lernenden.

Lernorte können um Netzwerkpartner (z. B. Testcenter) ergänzt werden, die effektiv und effizient Aufgaben übernehmen können und dadurch Entlastung für die Schulen bringen. Diese Freiräume sind nötig, um sich auf neue Inhalte zu spezialisieren. Im Mittelpunkt der Bildungsorganisation steht der Lernende, der seinen Kompetenzerwerb anhand eines breit definierten Kompetenzprofils in einem Berufsprofil mit personaler Lernbegleitung plant und umsetzt. Digitale Lehr-Lernformen richten sich auf integrierendes Lernen (z. B. Blended Learning, Flipped Classroom) aus, um den Kompetenzerwerb mit authentischen Aufgabenstellungen am Arbeitsplatz zu verknüpfen. Neue Organisationslogiken für das Gestalten von Bildungsprozessen liefern flexiblere Rahmenbedingungen, um die Individualisierung von Bildungsprozessen zu fördern.

Entwicklungsstufe II: Im digitalen Ökosystem

Da Plattformökonomien eine essentielle Voraussetzung für die Implementierung und kontinuierliche Verbesserung von KI (insbesondere das systematische Trainieren der KI-Systemen) darstellen, sind die Voraussetzungen und Rahmenbedingungen dafür auf Systemebene zu schaffen. Die gemeinsame Entwicklung mit Partnern und ein wechselseitiges Ermöglichen von neuen Aktivitäten sowie Bildungsleistungen in einem digitalen Ökosystem stehen dabei im Vordergrund.

Der Aufbau eines Ökosystems könnte neue Optionen bieten, mit zertifizierten Netzwerkpartnern tragfähige Geschäftsmodelle im Bildungswesen zu etablieren, wie z. B. zur Entwicklung von KI-Systemen sowie auch um Daten für Learning Analytics nutzbar machen zu können. Heute wird Learning Analytics eher zur Vorhersage im Rahmen eines Online-Kurses genutzt (wie hoch ist das Risiko, dass der Lernende den Kurs abbrechen wird?). Ein Lehrabbruch hat in der Regel weitreichende Folgen für den Lernenden. Inwieweit hier Voraussagen mit KI-basierten Verfahren getroffen werden können, untersucht beispielsweise DILLENBOURG (2017).

Im Rahmen von Smart Government eröffnen vernetzte Informations- und Kommunikationstechniken neue Möglichkeiten zur Analyse, Automation und Organisation von Bildungsprozessen (SEUFERT, 2018). So können KI-basierte Verfahren automatisch neue Bedarfe im Bereich der Berufsbildung aufspüren, indem sie Stellenangebote, gestellte Fragen sowie neue Normen und Regeln etc. analysieren (DILLENBOURG, 2017). Dazu wäre die den ‚Data Sciences' eigene Denkweise in das Management der Berufsbildung zu integrieren (ebenda).

6 Zusammenfassung und Ausblick

Für eine gelingende Berufsbildung ist eine erfolgreiche Kooperation der Lernorte Betriebe, Berufsschulen und überbetriebliche Ausbildungszentren nötig. Die Verbesserung der Lernortkooperation ist somit ein bildungspolitischer Dauerbrenner und es verwundert nicht, dass hierzu zahlreiche konzeptionelle und empirische Arbeiten vorliegen (vgl. Kapitel 3.2 und 3.3). Kernbotschaft kann sein, dass Lernortkooperation systematisch von den Lernenden her zu denken ist. Die Aktivitäten der Akteure der dualen beruflichen Bildung wären zu bündeln, um optimale Lernprozesse für die Lernenden zu ermöglichen. Wie derartige Lernprozesse aussehen können, die schulische und betriebliche Inhalte integrieren, zeigen STENSTRÖM und TYNJÄLÄ (2009). Hierzu sind Koordination und Abstimmung von Einzelaktivitäten nötig. Die gegenwärtige Ausprägung der Lernortkooperation kann als ein Gleichgewicht gesehen werden, in der die Grenzkosten der Koordination dem Grenznutzen durch höhere Lernerfolge entsprechen. Die fortgeschrittene Digitalisierung beeinflusst sowohl die Kosten- als auch auf Nutzenseite. Zum einen verringert die fortgeschrittene Digitalisierung die Koordinationskosten. Zum anderen steigt der Druck, flexiblere, agilere und stärker digitalisierte Arbeits- und Lernformen in der Berufsbildung zu entwickeln, zu erproben und deren Qualität nachhaltig sicherzustellen. M. a. W. der Nutzen aus kooperativem Lernen steigt.

Ein wesentliches Ziel des vorliegenden Beitrages ist es, Zukunftsszenarien und Gestaltungsoptionen zu entwerfen, um die digitale Transformation der Lernortkooperation in der Berufsbildung gestalten zu können. Im Vordergrund steht dabei, KI-basierte Prozesse vor allem für das personalisierte Lernen der Individuen im Rahmen der Lernortkooperation nutzen zu können. Darüber hinaus liegt der Konzeption zugrunde, die Organisationslogik der Lernortkooperation flexibler und digital, zunehmend auch KI-basiert, zu unterstützen. Die beiden Mesosysteme Schule und Lehrbetrieb sind somit in Kernprozessen vernetzt und Planungs-, Abstimmungs- und Entscheidungsprozesse können vereinfacht werden. Digitale Ökosysteme sind generell in der Bildung hierfür eine notwendige Voraussetzung. Zwar gibt es Forschung zu Ökosystemen im Bildungsbereich (s. zum Überblick SEUFERT, GUGGEMOS & MOSER, 2019), allerdings wären hier weitere Arbeiten nötig, um ein besseres Verständnis für deren Funktionsweise und Etablierung zu gewinnen. Gerade in föderalistisch geprägten Bildungssystemen scheinen sich derartige Ökosysteme schwer umsetzen zu lassen. Vielversprechend kann sein, sich an bereits existierenden Ökosystemen im Bildungsbereich zu orientieren. Hier kann die Initiative digital Israel ein Vorbild sein (IBL, 2018). Hierbei handelt es sich um eine open edX basierte Plattform, auf der Personen ihren lebenslangen Lernprozess gestalten können.

Wie aufgrund der skizzierten Überlegungen deutlich wird, ist Lernortkooperation nicht nur auf der Mikroebene zu verorten. Dennoch kann es sich anbieten, erste Use Cases zu entwickeln und zu evaluieren. Hierfür bieten sich Bereiche an, in denen keine hohen Anforderungen an die verfügbaren Datenmengen bestehen. Denkbar sind Anwendungen im Bereich Virtual Reality zur Simulation betrieblicher Realität als Kooperationsgrundlage und zur Verbindung von Arbeits- und Lernkontexten.

Strategisches Ziel der Use Cases und Evaluationen sollte es sein, das Verständnis für die digitale Transformation im Allgemeinen und die Rolle der Datennutzung im Speziellen zu fördern. Dabei wären auch kritische Aspekte zu berücksichtigen:

- *Privatsphäre*: Was geschieht mit den Daten der Lernenden?
- *Ethik*: Was sind mögliche Folgen einer Fehlinterpretation der Daten?
- *Normen*: Ist der soziale Vergleich bei Lernenden (z. B. Vergleich mit anderen Lernenden im gleichen Berufsfeld oder unterschiedlichen Berufsfeldern) opportun?
- *Zeitskala*: Sollen Analysen innerhalb oder außerhalb des Unterrichts durchgeführt werden?

Im Zusammenhang mit Learning Analytics sind somit die Gefahren und eine Reihe ethischer Fragestellungen zu berücksichtigen. Insbesondere das Anwendungsfeld der präskriptiven Analyse ist in diesem Zusammenhang als kritisch einzuschätzen. Datenverzerrungen und Fehlinterpretationen von Daten bergen Gefahren, weil diese zu suboptimalen Handlungen führen können. Daher bieten Reifegradmodelle für die digitale Transformation besonderes Potenzial, da evolutionär zunächst in Bereichen Erfahrungen gesammelt werden kann, die relativ unproblematisch erscheinen. Ein Beispiel wäre die Evaluation und Optimierung von Lernmaterialien mittels Learning Analytics.

Literatur

ACHTENHAGEN, F. / BENDORF, M. / WEBER, S. (2004). Lernortkooperation zwischen Wirklichkeit und „Vision". In D. EULER (Hrsg.). *Handbuch der Lernortkooperation, Band 1* (S. 77–101). Bielefeld: Bertelsmann.

AKKERMAN, S. F. / BAKKER, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169.

ALDOWAH, H. / AL-SAMARRAIE, H. / FAUZY, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics, 37*, 13–49.

APREA, C. / ARN, C. / BOLDRINI, E. / CATTANEO, A. / MOTTA, E. / SROKA, A. (2012). Digitale Technologien als Tools zur Förderung der Konnektivität des Lernens in Schule und Betrieb. In U. F. FASSHAUER / B. FÜRSTENAU / E. WUTTKE (Hrsg.). *Berufs- und wirtschaftspädagogische Analysen. Aktuelle Forschungen zur beruflichen Bildung* (S. 61–73). Opladen: Barbara Budrich.

APREA, C. / CATTANEO, A. P. / SAPPA, V. (2015). Mind the Gap: Boundary-Crossing an den Übergängen von informellem und formalem Lernen in der beruflichen Bildung. In G. NIEDERMAIR (Hrsg.). *Informelles Lernen: Annäherungen, Problemlagen, Forschungsbefunde Schriftenreihe für Berufs- und Betriebspädagogik* (S. 265–281). Linz: Trauner.

APREA, C., SAPPA, V. & TENBERG, R. (2020). Konnektivität und integrative Kompetenzentwicklung. Einleitung zum Themenheft. In: C. APREA / V. *Sappa* / R. TENBERG (Hrsg.). Konnektivität und lernortintegrierte Kompetenzentwicklung in der Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW), Sonderheft, 29,* 9–13.

APREA, C. / SAPPA, V. (2015). School-workplace connectivity: ein Instrument zur Analyse, Evaluation und Gestaltung von Bildungsplänen der Berufsbildung. *Berufsbildung in Wissenschaft und Praxis (BWP), 1,* 27–31.

APREA, C. / SAPPA, V. (2020). Individuelle Vorstellungen zum lernortübergreifenden Lernen und Lehren im schweizerischen Berufsbildungskontext. In: C. APREA / V. SAPPA / R. TENBERG (Hrsg.). Konnektivität und lernortintegrierte Kompetenzentwicklung in der Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW), Sonderheft, 29,* 165–184.

ARNOLD, R. (2020). Berufsbildung und Persönlichkeitsentwicklung in Zeiten disruptiver Innovationen. In R. ARNOLD / A. LIPSMEIER / M. ROHS (Hrsg.). Handbuch Berufsbildung (S. 28–39). Wiesbaden: Springer.

BAARTMAN, L. K. / DE BRUIJN, E. (2011). Integrating knowledge, skills and attitudes: conceptualizing learning processes toward vocational competence. *Educational Research Review, 6,* 125–134.

BECK, K. (1984). Zur Kritik des Lernortkonzeptes – Ein Plädoyer für die Verabschiedung einer untauglichen pädagogischen Idee. In W. GEORG (Hrsg.). *Schule und Berufsausbildung* (S. 247–262). Bielefeld: Bertelsmann.

BELLMAN, R. (1978). *An introduction to artificial intelligence: can computers think?* San Francisco: Boyd & Fraser.

BERGER, K. / WALDEN, G. (1995). Zur Praxis der Kooperation zwischen Schule und Betrieb – Ansätze zur Typisierung von Kooperationsaktivitäten und -verständnissen. In G. PÄTZOLD / G. WALDEN (Hrsg.). *Lernorte im dualen System der Berufsbildung* (S. 409–430). Bielefeld: Bertelsmann.

BILLETT, S. (2014). Integrating learning experiences across tertiary education and practice settings: a socio-personal account. *Educational Research Review, 12,* 1–13.

BRATER, M. (1992). Zwischen Anspruch und Wirklichkeit – Die Auswirkungen veränderter Rahmenbedingungen auf soziale Stellung, Funktion und Qualifizierungs-/Bildungskonzepte für das Ausbildungspersonal in der Industrie. In P. DEHNBOSTEL (Hrsg.). *Neue Technologien und berufliche Bildung. Modellhafte Entwicklung und theoretische Erkenntnisse. Berichte zur beruflichen Bildung* (S. 210–227). Heft 151. Bonn/ Berlin: Bundesinstitut für Berufsbildung (BIBB).

BUCKINGHAM SHUM, S. / DEAKIN CRICK, R. (2012). Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics. *Proceedings 2nd International Conference on Learning Analytics & Knowledge* (S. 324–335). ACM Press: New York

CARUSO, V. / CATTANEO, A. / GURTNER, J.-L. (2020). Exploring the Potential of Learning Documentation as a Boundary Object in the Swiss Vocational Education and Training System. In: C. APREA / V. SAPPA / R. TENBERG (Hrsg.). Konnektivität und lernortintegrierte Kompetenzentwicklung in der Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW), Sonderheft, 29,* 213–232.

CATTANEO, A. P. / APREA, C. (2018). Visual technologies to bridge the gap between school and workplace in vocational education. In D. IFENTHALER (Hrsg.). *Digital Workplace Learning: Bridging formal and informal learning with digital technologies* (S. 251–270). Cham: Springer.

DEHNBOSTEL, P. (2020). Lernorte und Lernortkooperation – Erweiterungen und Entgrenzungen nicht nur in digitalen Zeiten. *Berufsbildung in Wissenschaft und Praxis (BWP), Sonderheft Kooperation der Lernorte, 4,* 13–18.

DILLENBOURG, P. (2017). Digitale Möglichkeiten in der Berufsbildung. Transfer, Berufsbildung in Forschung und Praxis, *Schweizerische Gesellschaft für angewandte Berufsbildungsforschung (SGAB), 1.* https://www.sgab-srfp.ch/de/newsletter/digitale-moeglichkeiten-der-berufsbildung

DITTLER, U. (2017). Ein kurzer historischer Rückblick auf die bisherigen drei Wellen des E-Learning. In U. DITTLER (Hrsg.). *E-Learning 4.0. Mobile Learning, Lernen mit Smart Devices und Lernen in sozialen Netzwerken* (S. 5–42). Berlin: De Gruyter.

EDER, A. / KOSCHMANN, A. (2011). Die Rolle von Lernortkooperation bei der Umsetzung lernfeldorientierter Lehrpläne an berufsbildenden Schulen in Niedersachsen. *Berufs- und Wirtschaftspädagogik online – bwp@, 20.*

ERPENBECK, J. / VON ROSENSTIEL, L. (2007). *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis.* 2. Auflage. Stuttgart: Schäffer-Poeschel.

ERTL, H. (2020). Lernortkooperation über das traditionelle Verständnis hinausdenken. Sonderheft Kooperation der Lernorte. *Berufsbildung in Wissenschaft und Praxis (BWP), Sonderheft Kooperation der Lernorte, 4,* 3.

EULER, D. (2004). Lernortkooperation – eine unendliche Geschichte. In D. EULER (Hrsg.). *Handbuch der Lernortkooperation* (S. 12–24). Bielefeld: wbv.

EULER, D. (2015). Lernorte in der Berufsausbildung zwischen Potenzial und Realität. *Berufsbildung in Wissenschaft und Praxis (BWP), 1,* 2015, 6–9.

FASSHAUER, U. (2018). Lernortkooperation im Dualen System der Berufsausbildung – implizite Normalität und hoher Entwicklungsbedarf. In U. FASSHAUER, U. / R. ARNOLD / A. LIPSMEIER / M. ROHS (Hrsg.). *Handbuch Berufsbildung* (S. 25–38). Wiesbaden: Springer.

FERGUSON, R. / CLOW, D. / MACFADYEN, L. / ESSA, A. / DAWSON, S. / ALEXANDER, S. (2014). Setting learning analytics in context: overcoming the barriers to large-scale adoption. Journal of Learning Analytics, 1(3), 120–144.

GESSLER, M. (2017). The lack of collaboration between companies and schools in the German dual apprenticeship system: Historical background and recent data. *International Journal for Research in Vocational Education and Training,* 4(2), 164–195.

GIBSON, A. / KITTO, K. / WILLIS, J. (2014). A cognitive processing framework for learning analytics. Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (S. 212–216). ACM Press: New York.

GRELLER, W. / DRACHSLER, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. Educational Technology & Society, 15(3), 42–57.

GRIFFITHS, T. / GUILE, D. (2003). A Connective Model of Learning: The Implications for Work Process Knowledge. *European Educational Research Journal,* 2(1), 56–73.

GRIFFITHS, T. / GUILE, D. (2004). *Learning through work experience for the knowledge economy Issues for educational research and policy*. Cedefop Reference series; 48 Luxembourg: Office for Official Publications of the European Communities.

GUGGEMOS, J. / SEUFERT, S. / SONDEREGGER, S. (2020). Humanoid Robots in Higher Education: Evaluating the Acceptance of Pepper in the Context of an Academic Writing Course Using UTAUT. *British Journal of Educational Technology, 51*(5), 1864–1883.

HAUTZ, H. / OSTENDORF, A. (2020). Konnektivität im Betriebspraktikum als empirisches Phänomen und Ansatzpunkt einer Praktikumsdidaktik. In C. APREA / V. SAPPA / R. TENBERG (Hrsg.). Konnektivität und lernortintegrierte Kompetenzentwicklung in der Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW), Sonderheft 29,* 113–138.

HOFHUES, S. / SCHIEFNER-ROHS, M. (2017). Vom Labor zum medialen Bildungsraum. Hochschul- und Mediendidaktik nach Bologna. In C. IGEL (Hrsg.). Bildungsräume. *Proceedings der 25. Jahrestagung der Gesellschaft für Medien in der Wissenschaft,* 5. bis 8. September 2017 in Chemnitz (S. 32–43). Münster: Waxmann.

IBL (2018). *Campus-IL, Israel's National Open edX Platform, Consolidates Its Project*. https://iblnews.org/campus-il-israels-national-open-edx-platform-consolidates-its-project/.

IFENTHALER, D. / EGLOFFSTEIN, M. (2020). Development and implementation of a maturity model of digital transformation. *TechTrends, 64*, 302–309.

IFENTHALER, D. (2017). Designing Effective Digital Learning Environments: Toward Learning Analytics Design. Technology, Knowledge & Learning, 22, 401–404.

JUNG, T. (2010). Lernorte und Lernwege. *REPORT Zeitschrift für Weiterbildungsforschung, 2,* 5–7.

KERRES, M. / HEINEN, R. (2015). Open Informational Ecosystems: The Missing Link for Sharing Resources for Education. *International Review of Research in Open and Distributed Learning, 16* (1), 24–39.

KILKBRINK, V. / ENOCHSSON, A.-B. / SÖDERLIND, L. (2020). Digital Technology as Boundary Objects. Teachers' Experiences in Swedish Vocational Education. In C. APREA / V. SAPPA / R. TENBERG (Hrsg.). Konnektivität und lernortintegrierte Kompetenzentwicklung in der Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW), Sonderheft 29,* 233–245.

MAKRIDAKIS, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, 46–60.

OECD (2020). *Bildung auf einen Blick 2020: OECD-Indikatoren*, Bielefeld: wbv. https://www.oecd-ilibrary.org/education/bildung-auf-einen-blick_19991509.

OSTENDORF, A. (2019). Die Skill Ecosystem Perspektive als Denkrahmen zur Weiterentwicklung von Berufsbildungsstrukturen – eine Diskussion im Hinblick auf die digitale Transformation Titel des Beitrags. In: *bwp@ Spezial AT-2*: Beiträge zum 13. Österreichischen Wirtschaftspädagogik-Kongress, 1–14.

PÄTZOLD, G. (1995). Kooperation des Lehr- und Ausbildungspersonals in der beruflichen Bildung – Berufspädagogische Begründungen, Bilanz, Perspektiven. In G. PÄTZOLD / G. WALDEN (Hrsg.). *Lernorte im dualen System der Berufsbildung* (S. 143–166). Bielefeld: Bertelsmann.

PÄTZOLD, G. / WALDEN, G. (1995). Anspruch und Wirklichkeit der Lernortkooperation. In G. PÄTZOLD / G. WALDEN (Hrsg.*). Lernorte im dualen System der Berufsbildung* (S. 11–22). Bielefeld: Bertelsmann.

PÄTZOLD, G. (2003). *Lernfelder – Lernortkooperationen. Neugestaltung beruflicher Bildung 2. Auflage*. Bochum: Projekt Verlag.

PÄTZOLD, G. (1990). *Lernortkooperation. Impulse für die Zusammenarbeit in der beruflichen Bildung. Schriftenreihe Moderne Berufsbildung. Bd. 12*. Heidelberg: Sauer.

PETERSEN, A. W. / WEHMEIER, C. (2001). Evaluation der neuen IT-Berufe: Forschungskonzepte und Ergebnisse der bundesweiten BIBB-IT-Studie. In W. PETERSEN / F. RAUNER / F. STUBER (Hrsg.). *IT-gestützte Facharbeit. Gestaltungsorientierte Berufsbildung* (S. 283–310). Baden-Baden: NOMOS.

RAUNER, F. (1998). Von der dualen zu einer dual-kooperativen Berufsausbildung. In H. HOLZ / F. RAUNER / G. WALDEN (Hrsg.). *Ansätze und Beispiele der Lernortkooperation* (S. 175–195). Bielefeld: Bertelsmann.

RAUNER, F. (2017). Die Berufsschule – warum aus dem Juniorpartner ein Partner in der dualen Berufsausbildung werden sollte. *Online-Magazin für Arbeit – Bildung – Gesellschaft, 3*. http://denk-doch-mal.de/wp/felix-rauner-die-berufsschule-warum-aus-dem-juniorpartner-ein-partner-in-der-dualen-berufsausbildung-werden-sollte/?format=pdf.

RAUNER, F. / PIENING, D. (2015). Die Qualität der Lernortkooperation. A+B Forschungsbericht Nr. 20. Bremen: Forschungsnetzwerk Arbeit und Bildung. https://www.ibb.uni-bremen.de/files/upload/documents/publications/AB_20.pdf.

RENZ, A. / HILBIG, R. (2020). Prerequisites for artificial intelligence in further education: identification of drivers, barriers, and business models of educational technology companies. International Journal of Educational Technology in Higher Education, 17(1), 1–21.

RENZ, A., KRISHNARAJA, S. / GRONAU, E. (2020). Demystification of Artificial Intelligence in Education – How much AI is really in the Educational Technology? International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI), 2(1), 14–30.

ROHS, M. (2020). „Lernortkooperation findet in den Köpfen der Lernenden statt". Interview mit Prof. Dr. Matthias Rohs zum Wandel von Lernorten und deren Gestaltung. Sonderheft Kooperation der Lernorte. *Berufsbildung in Wissenschaft und Praxis (BWP), Sonderheft Kooperation der Lernorte, 4,* 8–10.

ROLL, M. / IFENTHALER, D. (2020). Lernortübergreifende Kompetenzentwicklung in der Industrie 4.0. Die Entwicklung digitaler Handlungskompetenz in der dualen Berufsausbildung aus der Ausbilderperspektive. In: C. APREA / V. SAPPA/ R. TENBERG (Hrsg.). Konnektivität und lernortintegrierte Kompetenzentwicklung in der Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW), Sonderheft, 29,* 185–209.

SAPPA, V. / APREA, C. (2014). Conceptions of Connectivity: How Swiss Teachers, Trainers and Apprentices Perceive Vocational Learning and Teaching Across Different Learning Sites. *Vocations and Learning, 7*(3), 263–287.

SCHALLMO, D. / RUSNJAK, A. / ANZENGRUBER, J. / WERANI, T. / JÜNGER, M. (2017). *Digitale Transformation von Geschäftsmodellen. Grundlagen, Instrumente und Best Practices*. Wiesbaden: Springer.

SCHNEID, R. (2018). Learning Factories in Vocational Schools. Challenges for Designing and Implementing Learning Factories at Vocational Schools. In D. IFENTHALER (Hrsg.). *Digital Workplace Learning – Bridging Formal and Informal Learning with Digital Technologies* (S. 271–289). New York: Springer.

SEUFERT, S. (2018). *Berufsbildung 2030. Flexibilisierung der Berufsbildung im Kontext fortgeschrittener Digitalisierung*. Bericht im Auftrag des Staatssekretariats für Bildung, Forschung und Innovation (SBFI) im Rahmen des Projekts „Berufsbildung 2030 – Vision und Strategische Leitlinien". Bern: SBFI.

SEUFERT, S. / GUGGEMOS, J. / MEIER, C. / HELFRITZ, K. (2020). *Trendstudie 2020. Auf dem Weg zur digital lernenden Organisation – Kompetenzen für die Personalentwicklung*. St. Gallen: Universität St. Gallen (scil) / Deutsche Gesellschaft für Personalentwicklung. https://www.scil.ch/produkt/scil-trendstudie-2020-auf-dem-weg-zur-digital-lernenden-organisation/.

SEUFERT, S. / GUGGEMOS, J. / SONDEREGGER, S. (2020). Digitale Transformation der Hochschullehre: Augmentationsstrategien für den Einsatz von Data Analytics und Künstlicher Intelligenz. *Zeitschrift für Hochschulentwicklung, 15*(1), 81–101.

SEUFERT, S. / GUGGEMOS, J. / MEIER, C. (2019). Shaping Digital Transformation: How Can HRD Competences be Conceptualised with a Focus on Augmentation? European Academy of Management, 26.–28. Juni, Lissabon.

SEUFERT, S. / GUGGEMOS, J. / MOSER, L. (2019). Digitale Transformation in Hochschulen: auf dem Weg zu offenen Ökosystemen. *Zeitschrift für Hochschulentwicklung, 14*(2), 85–107.

SEUFERT, S. / MEIER, C. / SÖLLNER, M. / RIETSCHE, R. (2019). A Pedagogical Perspective on Big Data and Learning Analytics: A Conceptual Model for Digital Learning Support. Technology Knowledge & Learning, 24, 599–619.

SEUFERT, S. / GUGGEMOS, J. / TARANTINI, E. (2018). Digitale Transformation in Schulen – Kompetenzanforderungen an Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung, 36*(2), 175–193.

STANFORD UNIVERSITY (2016). *Artificial intelligence and life in 2030. One hundred year study on artificial intelligence: report of the 2015–2016 study panel*. Stanford, CA. Standford University.

STAR, S. L. / GRIESEMIER, J. R. (1989). Institutional Ecology, ‚Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology. *Social Studies of Science, 19*(4), S. 387–420.

STENSTRÖM, M.-L. / TYNJÄLÄ, P. (2009). *Towards integration of work and learning. Strategies for connectivity and transformation*. Amsterdam: Springer.

SURMA, T. / KIRSCHNER, P. A. (2020). Technology enhanced distance learning should not forget how learning happens. *Computers in Human Behavior, 110*.

THORDSEN, T. / MURAWSKI, M. / BICK, M. (2020). How to Measure Digitalization? A Critical Evaluation of Digital Maturity Models. In M. HATTINGH / M. MATTHEE / H. SMUTS / I. PAPPAS / Y. DWIVEDI / M. MÄNTYMÄKI (Hrsg.). *Responsible Design, Implementation and Use of Information and Communication Technology*. I3E 2020. Lecture Notes in Computer Science, 12066. Cham: Springer.

TIPPELT, R. / REICH-CLASSEN, J. (2010). Lernorte: organisationale und lebensweltbezogenen Perspektiven. *REPORT – Zeitschrift für Weiterbildungsforschung, 2*, 11–21.

TULLY, C. (1994*). Lernen in der Informationsgesellschaft*. Informelle Bildung durch Computer und Medien. Opladen: Leske + Budrich.

TYNJÄLÄ, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*, 130–135.

TYNJÄLÄ, P. (2009). Connectivity and Transformation in Work-Related Learning – Theoretical Foundations. In M.-L. STENSTRÖM / P. TYNJÄLÄ (Eds.). *Towards integration of work and learning. Strategies for connectivity and transformation* (S. 11–33). Amsterdam: Springer.

TYNJÄLÄ, P. / VIRTANEN, A. / KLEMOLA, U. / KOSTIAINEN, E. / RAKSU-PUTTONEN, H. (2016). Developing social competence and other generic skills in teacher education: Applying the model of integrative pedagogy. *European Journal of Teacher Education, 39*(3), 368–387.

WAHLSTER, W. (2017). Künstliche Intelligenz als Treiber der zweiten Digitalisierungswelle. *IM+io Das Magazin für Innovation, Organisation und Management,* (2), 10–13.

WENNER, T. (2018). Entwicklung eines Instruments zur Erfassung der Wechselwirkung von Lernortkooperation und Ausbildungsqualität. *Journal of Technical Education, 6*(1), 223–237.

WETTSTEIN, E. / SCHMID, E. / GONON, P. (2014). *Berufsbildung in der Schweiz. Formen, Strukturen, Akzente*. Bern: hep.

ZAWACK-RICHTER, O. / MARIN, V. I. / BOND, M. / GOUVERNEUR, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(39), 1–27.

WINKLER, R. / SÖLLNER, M. (2018). Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analyses. *78th annual meeting of the Academy of Management Annual Meeting (AOM)*. Chicago, USA.

Zur Person: Sabine Seufert ist Direktorin des Instituts für Bildungsmanagement und Bildungstechnologien an der Universität St. Gallen (IBB-HSG), Schweiz. Ihre Forschung befasst sich mit der digitalen Transformation im Bildungsmanagement und Personalentwicklung, Künstliche Intelligenz und soziale Roboter im Bildungsbereich.

Prof. Dr. Sabine Seufert
Universität St. Gallen, Institut für Bildungsmanagement und Bildungstechnologien,
St. Jakob-Strasse 21, 9000 St. Gallen, Schweiz, sabine.seufert@unisg.ch

Zur Person: Josef Guggemos ist Assistenzprofessor für Bildungstechnologien und Informatisches Denken am Institut für Bildungsmanagement und Bildungstechnologien an der Universität St. Gallen (IBB-HSG), Schweiz. Seine Forschungsinteressen sind Informatisches Denken, Kompetenzen von Bildungsverantwortlichen im Kontext der digitalen Transformation und soziale Roboter im Bildungskontext.

Prof. Dr. Josef Guggemos
Universität St.Gallen, Institut für Bildungsmanagement und Bildungstechnologien,
St. Jakob-Strasse 21, 9000 St. Gallen, Schweiz, josef.guggemos@unisg.ch

Studie 8:

Guggemos, J. (2018). Analyse beruflicher Tätigkeitsfelder von Wirtschaftspädagogen/-innen anhand von Daten des Karriereportals XING. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 114(4), 551–577.

Franz Steiner Verlag

ZEITSCHRIFT FÜR BERUFS- UND WIRTSCHAFTSPÄDAGOGIK 114, 2018/4, 551–577

JOSEF GUGGEMOS

Analyse beruflicher Tätigkeitsfelder von Wirtschaftspädagogen/-innen anhand von Daten des Karriereportals XING

Occupational Fields of Business Education Scientists – Evidence from Social Career Network XING

KURZFASSUNG: Die Frage nach beruflichen Tätigkeitsfeldern von Wirtschaftspädagogen/-innen ist nicht einfach zu beantworten (SLOANE/TWARDY/BUSCHFELD 2004, S. 10). Ziel des vorliegenden Beitrags ist, Struktur und Niveau von Berufen, die Wirtschaftspädagogen/-innen außerhalb des Schuldiensts ausüben, zu untersuchen. Dazu dienen 2436 öffentlich verfügbare Profile auf dem Karriereportal XING. Die Berufsangaben wurden mit der Klassifikation der Berufe 2010 der Bundesagentur für Arbeit kodiert. Hinsichtlich der Struktur lassen sich mit 15 Berufsgruppen über 90 % der wirtschaftspädagogischen Tätigkeiten abdecken. Die beiden häufigsten sind mit über 40 % Anteil ‚Personalwesen und -dienstleistung' sowie ‚Unternehmensorganisation und -strategie'. Hinsichtlich des Niveaus der ausgeübten Berufe zeigt sich ein positiver Einfluss der Höhe des Studienabschlusses. Besonders ausgeprägt ist der Niveauzuwachs im Fall einer Promotion. Evidenz dafür, dass Wirtschaftspädagoginnen systematisch Berufe auf geringerem Niveau ausüben als Wirtschaftspädagogen, konnte nicht gefunden werden. 16 % der Wirtschaftspädagogen/-innen in der Stichprobe sind selbstständig tätig, am häufigsten als Unternehmensberater/-in.
Schlagworte: Wirtschaftspädagogik, Berufsstruktur, Berufsniveau, soziales berufliches Netzwerk XING, Webcrawling

ABSTRACT: It is not easy to answer the question about occupational fields of business education scientists (SLOANE/TWARDY/BUSCHFELD 2004, p. 10). The aim of this article is to examine the structure and level of occupations that business education scientists pursue outside the teaching profession. For this purpose, 2436 publicly available profiles on the social career network XING are utilized. The data were coded with the classification of occupations 2010 of the Federal Employment Agency in Germany. In terms of structure, 15 occupational groups cover more than 90 % of the business education scientists' occupations. The two most frequent, with a share of over 40 %, are ‚human resource management and services' and ‚company organization and strategy'. With regard to the occupational level, a positive influence of the level of the academic degree was found. The increase in occupational level is particularly pronounced in case of a PhD. There is no evidence that female business education scientists systematically pursue occupations at a lower level than their male counterparts. 16 % of the business education scientists in the sample are self-employed, most frequently as management consultants.

Keywords: business education science, occupational structure, occupational level, social occupational network XING, web crawling

1 Einführung

Die Frage, in welchen Berufen Wirtschaftspädagogen/-innen tätig sind, ist nicht einfach zu beantworten (SLOANE/TWARDY/BUSCHFELD 2004, S. 10). Die Polyvalenz des Studiengangs spiegelt sich im Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik wider. Dort sind als Tätigkeitsfelder für Wirtschaftspädagogen/-innen genannt (SEKTION BWP DER DGFE 2014, S. 6): Berufliches Schulwesen, Betriebliches Bildungs- und Personalwesen, Berufliche Weiterbildung, Bildungsverwaltung, Bildungsmanagement und Bildungspolitik, Bildungsberatung, wirtschaftswissenschaftliche Tätigkeitsfelder sowie Wissenschaft. Um diese Breite abzudecken, werden Wirtschaftspädagogen/-innen ausgebildet in Betriebs- und Volkswirtschaftslehre, Wirtschaftsdidaktik sowie relevanten Bezugswissenschaften wie Statistik, Recht oder Wirtschaftsinformatik (KULTUSMINISTERKONFERENZ 2017, S. 74 f.; GRAMLINGER/SCHLÖGL/STOCK 2008; IWP 2018).

Welche Berufe Wirtschaftspädagogen/-innen neben der unmittelbar evidenten Tätigkeit an beruflichen Schulen tatsächlich ausüben, ist für Deutschland und die Schweiz weitgehend unerforscht. Für Österreich liegen Absolventenbefragungen für die Standorte Graz und Wien vor (z. B. ZEHETNER/STOCK/SLEPCEVIC-ZACH 2016; STOCK/FERNANDEZ/SCHELCH 2007; HAUER/STOCK 2007). Erkenntnisinteresse dieser Studien sind Karriereverläufe, Zufriedenheit mit der beruflichen Tätigkeit, Nutzen des Studiums für die betriebliche Praxis und Gründe für die Wahl des Studiums. Die Arbeiten liefern für die beiden Standorte einen vertieften Einblick. Für einen Gesamtüberblick der wirtschaftspädagogischen Tätigkeitsfelder in den DACH-Ländern sind die Studien jedoch eher nicht geeignet. Darüber hinaus adressieren diese Studien auch eher weniger das Niveau, auf dem Wirtschaftspädagogen/-innen Tätigkeiten ausüben, beispielsweise ob Promovierte im Vergleich zu Masterabsolventen/-innen in Berufen auf höherem Niveau beschäftigt sind. Die Untersuchung wirtschaftspädagogischer Tätigkeitsfelder ist in mehrerlei Hinsicht eine relevante Forschungslücke, wie nachfolgend aufgezeigt wird.

Bei der Wahl eines Studiums vermissen Abiturienten/-innen Informationen und Aussagen dazu, welche Tätigkeitsfelder sich hinter dem Studium verbergen (HACHMEISTER/HARDE/LANGER 2007, S. 38). Für den Studiengang Wirtschaftspädagogik sind diese, abgesehen von den Beschäftigungsmöglichkeiten im staatlichen Schuldienst, nicht unmittelbar ersichtlich. Vor diesem Hintergrund würden angehende Studierende von Wissen zu ausgeübten Tätigkeiten von Wirtschaftspädagogen/-innen profitieren. Wird als Gegenstand einer Berufsberatung das Aufzeigen von Alternativen gesehen (GATI/ASHER 2001), können auch Berufs- oder Karriereberater/-innen dieses Wissen nutzen. In den Beschreibungen von Studiengängen finden sich zwar Angaben zu wirtschaftspädagogischen Tätigkeitsfeldern außerhalb des Schuldiensts (z. B. ZLATKIN-TROITSCHANSKAIA/BREUER 2010, S. 132), eine empirische Untermauerung dieser

Angaben fehlt aber regelmäßig. Das gilt auch für Portale der Bundesagentur für Arbeit in Deutschland oder der Wirtschaftskammer in Österreich (z. B. BUNDESAGENTUR FÜR ARBEIT 2006; INSTITUT FÜR BILDUNGSFORSCHUNG DER WIRTSCHAFT o. J.).

Das Basiscurriculum (SEKTION BWP DER DGFE 2014, S. 11) sieht im Inhaltsbereich „Professionalisierung" für den Bachelor „Tätigkeitsfelder für Berufs- und Wirtschaftspädagogen/innen" vor. Dieser Bereich kann, wie beispielsweise an der LMU München, über eine Berufsfelderkundung abgedeckt werden. Um eine Auswahl relevanter Berufsfelder zu treffen, kann Wissen zu den tatsächlich gewählten Berufsfeldern nützlich sein. Weiterhin ist für den Bachelor die „Theoriegeleitete Bearbeitung praxisnaher Fragestellungen und Probleme" vorgesehen. Auch hier können Verantwortliche für Lehrveranstaltungen bei der Auswahl der Inhalte von einer belastbaren Datengrundlage zu den tatsächlich ausgeübten Berufen profitieren.

Ein Ziel der Bologna-Reform war, den Bachelor als Regelabschluss zu etablieren. Dieses Ziel scheint zumindest für Deutschland verfehlt worden zu sein. Hier münden zwei Drittel der Bachelorabsolventen nach dem Abschluss in ein Masterstudium ein (CHRISTOPH/LEBER/STÜBER 2017). Wie das Verhältnis von Master- zu Bachelorabschlüssen bei Wirtschaftspädagogen/-innen insgesamt ausgeprägt ist, ist eine unbeantwortete Frage. Das Anstreben eines Masterabschlusses ist mit Erwartungen verbunden. Personen könnten sich dafür interessieren, ob mit einem höheren Abschluss auch ein Potential für eine Beschäftigung auf höherem Niveau verbunden ist. Beispielsweise könnten sich Wirtschaftspädagogen/-innen mit Bachelorabschluss fragen, ob mit Masterabschluss im Vergleich zum Bachelor im Durchschnitt Berufe auf höherem Niveau ausgeübt werden.

Um zur Schließung der Forschungslücke beizutragen, untersucht der vorliegende Beitrag systematisch die von Wirtschaftspädagogen/-innen gewählten Berufe. Dazu dienen Daten des Karriereportals XING. Als Wirtschaftspädagoge/-in wird verstanden, wer mindestens einen (Universitäts-)Abschluss in Wirtschaftspädagogik erworben hat. Das Konstrukt ‚Beruf' wird primär unter Rückgriff auf die Klassifikation der Berufe 2010 (KldB 2010) der Bundesagentur für Arbeit in Deutschland abgegrenzt. Demnach lässt sich ein Beruf über zwei Dimensionen beschreiben: Berufsfachlichkeit und Anforderungsniveau. Als Berufsfachlichkeit wird verstanden „ein auf berufliche Inhalte bezogenes Bündel von Fachkompetenzen" (BUNDESAGENTUR FÜR ARBEIT 2011, S. 26). Allerdings scheint es verkürzt, Berufe inhaltlich auf diesen Aspekt zu beschränken. Vielmehr zeichnen sich diese auch durch spezifische Interessenprofile aus (BERGMANN/EDER 2005; WEBER/GUGGEMOS 2018). Berufsfachlichkeit und Interessenprofil werden unter dem Überbegriff Berufsstruktur diskutiert. Die Forschungsfragen sind vor diesem Hintergrund:
1) Wie ist die Berufsstruktur bei Wirtschaftspädagogen/-innen ausgeprägt?
2) Auf welchem Niveau üben Wirtschaftspädagogen/-innen Berufe aus?
3) Welche Faktoren erhöhen die Wahrscheinlichkeit einer Selbstständigkeit?

Der vorliegende Beitrag soll über eine reine Beschreibung hinausgehen. Zu diesem Zweck werden im nachfolgenden Abschnitt Hypothesen zu möglichen Unterschieden bei der Berufsstruktur sowie zu Einflüssen auf das Berufsniveau und die Wahrscheinlichkeit einer Selbstständigkeit gebildet.

2 Theoretischer Hintergrund und Hypothesen

Berufswahltheorien zielen darauf ab, Berufswahlentscheidungen zu erklären. Nachfolgend werden wichtige Berufswahltheorien mit Bezug zur Forschungsfrage dargestellt. Dabei handelt es sich um eine Betrachtung aus der Perspektive des Individuums. Weil diese nur eine Seite der Medaille darstellt, findet zusätzlich die Arbeitsmarktperspektive bzw. die Arbeitgeberseite Berücksichtigung.

Nach GOTTFREDSON (1981, 2002) findet bereits in der Kindheit und Jugend ein Eingrenzungsprozess statt. Als mögliche Alternativen werden Berufe ausgeschlossen, die nicht zum eigenen Geschlecht, den eigenen Fähigkeiten und zum persönlichen Interesse passen. Bei der tatsächlichen Berufswahl sind regelmäßig Kompromisse einzugehen. Personen sind dabei am ehesten zu Abstrichen beim persönlichen Interesse bereit, dann beim Prestige und schließlich beim Geschlechtstyp des Berufs. Der Geschlechtstyp ist eine Einschätzung, wie maskulin oder feminin ein Beruf ist. Das Prestige spiegelt den kognitiven Anspruch an einen Beruf und damit das Niveau wider. Das Interesse erfasst GOTTFREDSON im Rückgriff auf die Berufswahltheorie von HOLLAND (1997). Es wird demnach über sechs Profile beschrieben: ‚realistic', ‚investigative', ‚artistic', ‚social', ‚enterprising' und ‚conventional' (RIASEC). Diese sind bei Personen unterschiedlich stark ausgeprägt. Einzelne Berufe wiederum lassen sich ebenfalls über die RIASEC-Ausprägungen charakterisieren. Hohe Arbeitszufriedenheit stellt sich potentiell ein, wenn eine gute Übereinstimmung zwischen persönlichen Interessen und Berufsanforderungen hinsichtlich der RIASEC-Merkmale gegeben ist.

Weil der Geschlechtstyp vor Prestige und Interesse zur Wahrung des Selbstkonzepts die bedeutendste Rolle spielt, ist ein Unterschied zwischen Frauen und Männern hinsichtlich der Berufsstruktur zu vermuten. Wirtschaftspädagoginnen dürften somit tendenziell typisch männliche Berufe ausschließen, Wirtschaftspädagogen typisch weibliche. Eine Ungleichverteilung von Männern und Frauen auf Berufe zeigt sich empirisch über viele Länder hinweg (CHARLES/BRADLEY 2009). Auch zwischen dem Interessenprofil und dem Geschlecht besteht nach GOTTFREDSON ein Zusammenhang. Als eher mit einem männlichen Geschlechtstyp verbunden gelten die Profile ‚realistic', ‚enterprising' und ‚investigative', als eher mit einem weiblichen ‚conventional' (GOTTFREDSON 1981, S. 553). Zwar kann konzeptionell von einem tatsächlich ausgeübten Beruf nicht auf das Interesse geschlossen werden. Personen könnten aufgrund der Rahmenbedingungen in für sie weniger interessanten Berufen tätig sein. Allerdings zeigt sich empirisch ein sehr enger Zusammenhang zwischen Geschlecht und Interessenprofil des ausgeübten Berufs (GOTTFREDSON 1981, S. 553). Vermutet wird somit auch bei Wirtschaftspädagogen/-innen ein Zusammenhang zwischen den RIASEC-Merkmalen der ausgeübten Berufe und dem Geschlecht. Bei den von Männern gewählten Berufen wird eine höhere Ausprägung der Profile ‚realistic', ‚enterprising' und ‚investigative', bei Frauen des Profils ‚conventional' erwartet.

Personen streben einen Beruf an, der im Abgleich mit dem Selbstkonzept ein zumindest akzeptables Prestige verspricht. Das Prestige eines Berufs verhält sich dabei proportional zu den kognitiven Anforderungen an den Beruf (GOTTFREDSON 2002,

S. 88). Personen mit höherem Selbstkonzept über die eigene Leistungsfähigkeit streben höhere Abschlüsse an (GOTTFREDSON 2002, S. 98). Das Selbstkonzept wird durch Erfahrungen des Individuums geprägt (SHAVELSON/HUBNER/STANTON 1976). Im beruflichen Kontext kann die Berufserfahrung zur Steigerung des Selbstkonzepts beitragen (MORTIMER/LORENCE 1979, S. 308 f.). Eine höhere berufliche Aspiration mit zunehmender Erfahrung scheint damit wahrscheinlich. Bei der Frage nach Geschlechterunterschieden hinsichtlich des Prestiges zeigt sich, dass es mehr Berufe mit sehr hohem Prestige mit männlichem Geschlechtstyp gibt (GOTTFREDSON 2002, S. 101). In Deutschland sind Frauen tatsächlich seltener in höheren beruflichen Positionen und Führungsfunktionen tätig (ACHATZ 2005, S. 290 f.). Insgesamt betrachtet zeigen sich nach GOTTFREDSON bei berufstätigen Männern und Frauen allerdings keine Unterschiede hinsichtlich des Prestiges der ausgeübten Berufe (1981, S. 553).

Zur Ableitung von Hypothesen über das Niveau der Beschäftigung sind Berufswahltheorien alleine nur bedingt geeignet. Ob sich das Angestrebte realisieren lässt, lässt sich ohne Berücksichtigung der Arbeitgeberseite schwer beantworten. Arbeitgeber sind aufgrund der Informationsasymmetrie über die (kognitive) Leistungsfähigkeit potentieller Arbeitnehmer bei der Personalauswahl auf Signale angewiesen. Glaubwürdige Signale in diesem Kontext können Studienabschlüsse sein (SPENCE 1973). Arbeitgeber werden mit steigendem Anspruch an die (kognitive) Leistungsfähigkeit des Berufs zunehmend höhere Abschlüsse einfordern.

Der Humankapitaltheorie nach MINCER folgend hat Berufserfahrung einen positiven Einfluss auf die Produktivität (MINCER 1974, S. 64 ff.). Zu erwarten ist ein konkaver Zusammenhang zwischen Produktivität und Berufserfahrung (MINCER 1974, S. 84). Für einen solchen Zusammenhang gibt es auch empirische Evidenz (MARANTO/RODGERS 1984). Auch bei Wirtschaftspädagogen/-innen wäre ein positiver, unterlinearer Einfluss der Berufserfahrung auf das Prestige des Berufs zu erwarten.

Wirtschaftspädagogen/-innen sind nicht nur abhängig beschäftigt, sondern auch selbstständig tätig. Das Interessenprofil ‚enterprising' entspricht eher einem männlichen Geschlechtstyp (GOTTFREDSON 1981, S. 553). Auch empirisch zeigt sich in internationalen Studien ein Einfluss des Geschlechts auf die Entscheidung für eine Selbstständigkeit (KOELLINGER/MINNITI/SCHADE 2013). Frauen sind auch nach Berücksichtigung von Kontrollvariablen seltener als Männer selbstständig tätig. Faktoren, die hierfür eine Rolle spielen, sind eine höhere Angst vor einem Scheitern und ein geringeres unternehmerisches Selbstkonzept bei Frauen. Somit wäre auch bei Wirtschaftspädagogen/-innen ein negativer Einfluss des Geschlechts ‚weiblich' auf die Wahrscheinlichkeit einer Selbstständigkeit zu vermuten. Aufgrund der Abhängigkeit des Selbstkonzepts von Erfahrungen ist ferner ein positiver Einfluss von Studienabschlüssen und Berufserfahrung auf die Wahrscheinlichkeit einer Selbstständigkeit zu erwarten. Empirisch zeigt sich für Deutschland, dass Personen regelmäßig zunächst einer abhängigen Beschäftigung nachgehen und sich erst später im Verlauf der beruflichen Karriere möglicherweise für eine Selbstständigkeit entscheiden. Darüber hinaus lässt sich ein positiver Zusammenhang zwischen der Höhe der Abschlüsse und der Wahrscheinlichkeit einer Selbstständigkeit feststellen (MAI/MARDER-PUCH 2013).

Zusammenfassend werden folgende Hypothesen für Wirtschaftspädagogen/-innen aufgestellt:

Abhängig Beschäftigte

H1a Es besteht ein Zusammenhang zwischen der Berufsfachlichkeit und dem Geschlecht.

H1b Es besteht ein Zusammenhang zwischen der Ausprägung der RIASEC-Merkmale der ausgeübten Berufe und dem Geschlecht: Männer weisen eine höhere Ausprägung bei ‚realistic', ‚enterprising' und ‚investigative' auf, Frauen bei ‚conventional'.

H2a Es besteht ein positiver Einfluss der Höhe des Studienabschlusses auf das Niveau des ausgeübten Berufs.

H2b Es besteht ein konkaver Zusammenhang zwischen Berufserfahrung und dem Niveau des ausgeübten Berufs.

H2c Das Geschlecht ‚weiblich' hat einen negativen Einfluss auf das Niveau des ausgeübten Berufs.

Selbstständige

H3a Es besteht ein Zusammenhang zwischen der Berufsfachlichkeit und dem Geschlecht.

H3b Es besteht ein positiver Einfluss der Höhe des Studienabschlusses auf die Wahrscheinlichkeit einer Selbstständigkeit.

H3c Es besteht ein konkaver Zusammenhang zwischen Berufserfahrung und der Wahrscheinlichkeit einer Selbstständigkeit.

H3d Das Geschlecht ‚weiblich' hat einen negativen Einfluss auf die Wahrscheinlichkeit einer Selbstständigkeit.

3 Stichprobe und Methode

3.1 Nutzung von Daten des Karriereportals XING

Zur Gewinnung von Daten für die Beantwortung der Forschungsfragen dient XING als das marktführende soziale Netzwerk für berufliche Kontakte im deutschsprachigen Raum (XING AG 2017, S. 2; BÄRMANN 2017, S. 10; SCHWÄR 2018). Im Vergleich zu privaten sozialen Netzwerken (z. B. Facebook), steht das berufliche Profil im Mittelpunkt. Personen können u. a. ihre bisher ausgeübten Tätigkeiten und erworbenen Abschlüsse angeben. Die Nutzungsbedingungen schreiben dabei vor: „Der Nutzer sichert zu, dass alle von ihm angegebenen Daten wahr und vollständig sind. Der Nutzer ist verpflichtet, die Daten hinsichtlich aller von ihm genutzten Anwendungen während der gesamten Vertragslaufzeit wahr und vollständig zu halten".[1] Unabhängig davon ist eine

1 S. https://www.xing.com/terms, Punkt 2.3.

sorgsame Pflege des Profils für einen professionellen beruflichen Auftritt ratsam (BÄRMANN 2017). Vor diesem Hintergrund scheinen die XING-Daten zur Erschließung wirtschaftspädagogischer Tätigkeitsfelder grundsätzlich geeignet. Das gilt insbesondere, weil der Studiengang in den DACH-Ländern angeboten wird.

Ausgangspunkt bilden alle 6.8 Millionen öffentlich verfügbaren XING-Profile. Diese wurden im November 2016 mithilfe der Statistiksoftware R (Version 3.3.1) automatisiert aufgerufen und dabei folgende Daten extrahiert (Webcrawling):

- Vor- und Nachname;
- aktuell ausgeübter Beruf, seit wann dieser ausgeübt wird und bei welchem Arbeitgeber;
- die letzten beiden erworbenen Abschlüsse mit jeweiliger Hochschule sowie Beginn und Ende des Ausbildungsgangs;
- Sprachkenntnisse mit Selbsteinschätzung dazu.

3.2 Aufbereitung des Datensatzes

2873 Personen verfügen nach eigenen Angaben über einen Abschluss in Wirtschaftspädagogik bzw. als Diplom-Handelslehrer/-in. Diese bilden die Basis für die weiteren Auswertungen und Aufbereitungen des Datensatzes. Die große Herausforderung dabei ist die mit Ausnahme der Ausprägung der Sprachkenntnisse fehlende Standardisierung der Benutzereingaben. Aus diesem Grund ist es nötig, die angegebenen Berufe inhaltsanalytisch auszuwerten. Hierzu dient die KldB 2010 als deduktives Kategoriensystem (BUNDESAGENTUR FÜR ARBEIT 2011; PAULUS/MATTHES 2013; ZÜLL 2016). Drei Gründe waren hierfür ausschlaggebend: (1) Die KldB berücksichtigt explizit die deutschsprachige Arbeitsmarktstruktur mit seinem spezifischen Berufskonzept; (2) In der KldB existiert eine abschließende Liste an Berufsbenennungen, die das Kodieren erleichtert; (3) Eine Transformation von KldB 2010 in die ‚International Standard Classification of Occupations 2008' (ISCO-08) ist über Umsteigeschlüssel möglich (BUNDESAGENTUR FÜR ARBEIT 2013). Auf diese Weise ist auch eine Verknüpfung mit der umfangreichen O*Net-Datenbank (PETERSON u. a. 2001) möglich. Dazu werden über die ISCO-08-Codes und Umsteigeschlüssel (BUREAU OF LABOR STATISTICS 2012) die im O*Net verwendeten ‚Standard Occupational Classification Codes' (SOC-Codes) ermittelt (O*NET 2017).

Die KldB 2010 enthält 23993 Berufsbenennungen, denen jeweils ein fünfstelliger Code zugeordnet ist. Die ersten vier Stellen der KldB geben Aufschluss über die Berufsfachlichkeit: Die erste Stelle determiniert den Berufsbereich, die zweite die Berufshauptgruppe, die dritte die Berufsgruppe und die vierte die Berufsuntergruppe. Die fünfte Stelle umfasst vier Ausprägungen und gibt Aufschluss über das Niveau (BUNDESAGENTUR FÜR ARBEIT 2011, S. 40):

- Level 1: „Helfer, Beamte einfacher Dienst, einjährige Berufsausbildungen";
- Level 2: „Fachkräfte, Beamte mittlerer Dienst";

- Level 3: „Meister, Techniker u. a., Kaufmännische Fortbildungen u. a., Beamte gehobener Dienst, Bachelorstudiengänge“ und
- Level 4: „Studienberufe (mind. vierjährig), Beamte höherer Dienst“.

Die Kodierung[2] der angegebenen Berufe mithilfe der KldB ist reliabel (LANDIS/KOCH 1977, S. 165). Eine vom Autor durchgeführte Kodierung von 100 zufällig ausgewählten Berufsnennungen ergab 92 Übereinstimmungen mit der initialen Kodierung. Cohen's Kappa beträgt .914. Bei der Inspektion der abweichenden Kodierungen zeigten sich keine systematischen Verzerrungen. Vielmehr handelte es sich um tatsächlich schwer zuordenbare Angaben wie „Head of Research“ (die Person arbeitet in einer Personalberatung) oder die abweichende Kodierung des Niveaus, beispielsweise „Senior Financial Accounting“ Stufe 3 vs. Stufe 4. Vor dem Hintergrund der berichteten Reliabilitätswerte diente die initiale Kodierung für alle weiteren Auswertungen.

Das Geschlecht der Personen wurde anhand einer online verfügbaren Liste männlicher und weiblicher Vornamen[3] automatisiert ermittelt. Die Hochschule, an der der Abschluss erworben wurde, konnte lediglich halbautomatisiert ausgewertet werden. Ca. 15 % der Angaben mussten manuell kodiert werden, z. B. ‚GU FFM‘ für die ‚Johann Wolfgang Goethe Universität Frankfurt‘.

Anschließend wurden anhand der KldB-Codes 403 Lehrkräfte (m = 43.92 %/w = 56.08 %) bzw. 34 Schulleitungspersonen (m = 67.65 %/w = 32.35 %) ausgeschlossen. Das sind 15.21 % der initialen Datenbasis. Der Grund für den Ausschluss: Wirtschaftspädagogen/-innen im Schuldienst sind, unabhängig von der Berufserfahrung, stets dem Level 4 der KldB zuzuordnen. Außerdem ist eine solche Tätigkeit mit einem Bachelorabschluss grundsätzlich nicht möglich. Untersuchungen des Einflusses des Abschlusses oder der Erfahrung auf das Niveau sind aufgrund fehlender Varianz des Niveaus wenig sinnvoll.

Eine Sonderrolle nehmen Selbstständige ein. Darunter werden Personen verstanden, die sich auf XING als ‚Unternehmer/-in‘, ‚Freiberufler/-in‘ oder ‚selbstständig‘ ausgeben. Selbstständigkeit hat neben dem konkreten Fachinhalt stets auch eine unternehmerische Komponente. Die 378 Selbstständigen werden daher separat analysiert.

3.3 Beschreibung der Stichprobe

Abhängig Beschäftigte

Von den 2058 abhängig Beschäftigten sind 1194 (58.02 %) Frauen. Als höchsten Abschluss erwarben einen Bachelor 7.63 % (m = 40.76 %/w = 59.24 %), ein Diplom 44.56 % (m = 49.29 %/w = 50.71 %), einen Magister 22.06 % (m = 33.26 %/w = 66.74 %), einen Master 21.62 % (m = 33.49 %/w = 66.51 %) und eine Promotion 4.13 % (m = 58.16/w = 41.84 %). Diese Abschlüsse liegen im Durchschnitt 10.00 Jahre

2 Dank gilt SABRINA GRÜNING, KATRIN HARTMANNSEGGER und THEA SCHNEIDER für die Kodierung.
3 http://www.beliebte-vornamen.de/28071-derzeit-lebende-bevoelkerung.htm.

(m = 12.08/w = 8.50) zurück (SD = 6.79 [m = 7.78/w = 5.50]). Am häufigsten erwarben die Personen den höchsten Wirtschaftspädagogikabschluss in Wien (12.39 %), Linz (8.79 %), Mannheim (6.96 %), Graz (6.41 %), München (LMU) (5.88 %), Innsbruck (4.81 %), Köln (4.42 %) und Göttingen (4.37 %). Insgesamt erhielten den Abschluss in Österreich 32.40 % (m = 31.33 %/w = 68.67 %), in der Schweiz 3.60 % (m = 72.97 %/w = 27.03 %) und in Deutschland 64.00 % (m = 45.63 %/w = 54.37 %). Dass es sich bei Wirtschaftspädagogik um ein deutschsprachiges Phänomen handelt, zeigen auch die Angaben zu den Sprachkenntnissen. Lediglich 15 Personen (0.73 %) geben an, Deutsch sei nicht ihre Muttersprache. Die am häufigsten genannten Fremdsprachen sind Englisch (90.33 %), Französisch (43.65 %), Spanisch (18.15 %), Italienisch (11.69 %) und Russisch (5.10 %).

Selbstständige
Von den 378 Selbstständigen sind 161 (42.59 %) Frauen. Als höchsten Abschluss haben einen Bachelor 1.59 % (m = 83.33 %/w = 16.67 %), ein Diplom 51.06 % (m = 58.19 %/w = 41.81 %), einen Magister 23.81 % (m = 45.56 %/w = 54.44 %), einen Master 11.90 % (m = 57.78 %/w = 42.22 %) und eine Promotion 11.64 % (m = 75.00 %/w = 25.00 %). Diese Abschlüsse liegen im Durchschnitt 17.65 Jahre (m = 18.78/w = 16.11) zurück (SD = 9.21 [m = 10.43/w = 7.02]). Am häufigsten erwarben die Selbstständigen den höchsten Wirtschaftspädagogikabschluss in Wien (14.55 %), München (LMU) (6.88 %), Hamburg (6.35 %), Linz (6.08 %), Graz (5.56 %), Frankfurt (5.29 %), Berlin (5.29 %), Mannheim (5.03 %) und Köln (5.03 %). Insgesamt erhielten den Abschluss in Österreich 30.95 % (m = 47.01 %/w = 52.99 %), in der Schweiz 2.65 % (m = 70.00 %/w = 30.00 %) und in Deutschland 66.40 % (m = 61.75 %/w = 38.25 %).

3.4 Methode

3.4.1 Operationalisierung der Variablen

Als Operationalisierung der Berufsfachlichkeit dient die Berufsgruppe, d. h. die ersten drei Stellen des KldB-Codes. Die Berufshauptgruppe (zwei Stellen) wäre für diesen Zweck zu grob. So verwundert es nicht, dass Wirtschaftspädagogen/-innen außerhalb des Schuldiensts primär in Unternehmensführung und -organisation arbeiten (48.3 %). Auf der anderen Seite wäre eine Analyse auf Ebene der Berufsuntergruppe (vier Stellen) zu detailliert. Die Berufsgruppe bildet einen guten Kompromiss der Aggregation. Auch andere Studien verwenden diese, um berufliche Tätigkeiten zu erfassen (Brücker u. a. 2012, S. 162).

Die Interessenprofile der ausgeübten Berufe werden über die RIASEC-Merkmale der ausgeübten Berufe erfasst. Diese können auf indirektem Weg ermittelt werden. Über die KldB-Codes ist eine Verknüpfung mit der O*Net-Datenbank möglich (O*Net 2017). Jede Charakterisierung einer Tätigkeit im O*Net enthält die Stärke der Ausprägung der sechs RIASEC-Merkmale auf einer Skala von 1 bis 7. Beispielsweise werden der Tätigkeit eines ‚Chief Executives' folgende Werte zugeordnet: 1.33 ‚realistic', 2.00 ‚investigative', 2.67 ‚artistic', 3.67 ‚social', 7.00 ‚enterprising' und 5.33 ‚conventional'.

Aufgrund der KldB-Kodierung liegt ein Maß für das Niveau des ausgeübten Berufs unmittelbar vor. Die fünfte Stelle der KldB weist aber lediglich vier Abstufungen auf. Außerdem wäre vor dem theoretischen Hintergrund auf das Prestige des Berufs als Maß für das Niveau abzuzielen. GOTTFREDSON (2002, S. 91) verweist für die Erfassung des Prestiges auf TREIMAN. Die Variable ‚Prestige' leitet sich dabei aus einer Befragung in 55 Ländern zum Prestige der Berufe nach ISCO ab (GANZEBOOM/TREIMAN 1996). Eine Datenbank mit einer Zuordnung von Prestigewerten zu ISCO-08 Codes ist frei verfügbar (GANZEBOOM/TREIMAN 2012). Unter allen Berufen am wenigsten prestigeträchtig ist der Beruf des Schuhputzers (12.00), am höchsten der des Universitätsprofessors (78.16). Die Variable ‚Prestige' wird logarithmiert, um die Schiefe der Verteilung in der Stichprobe abzumildern. Die Korrelation (Kendalls tau) zwischen KldB-Level und Prestige beträgt .548.

Zur Operationalisierung der Höhe des Abschlusses ist die Bildung einer Rangfolge nötig. Diese ist nicht für alle Fälle möglich. Beispielsweise ist nicht klar, ob ein Diplom- oder ein Masterabschluss als höher einzuschätzen ist. Folgende Abstufung scheint aber angemessen: Die Promotion ist der höchste Abschluss, der Bachelor der niedrigste. Die weiteren Abschlüsse bilden die Mittelgruppe, in der keine weitere Differenzierung vorgenommen wird.

Als Operationalisierung der Erfahrung fungiert die Anzahl der Jahre, die seit dem höchsten erworbenen Studienabschluss vergangen sind. Zur Erleichterung der Interpretation der Regressionsparameter wird diese Variable mit dem Mittelwert zentriert.

Als Kontrollvariable aufgenommen wird bei der Analyse des Niveaus das Land, in dem der Abschluss erworben wurde: Österreich, Schweiz und Deutschland. Dadurch sollen Niveauunterschiede in den Ländern Berücksichtigung finden.

3.4.2 Prüfung der Hypothesen

Hypothese 1

Zur Prüfung des Zusammenhangs zwischen der Berufsgruppe und dem Geschlecht dient ein χ^2-Test. Die p-Werte werden mithilfe von Bootstrapping (n = 5000) ermittelt, weil die Annahmen des parametrischen χ^2-Tests nicht immer erfüllt sind. Die Effektstärke der Zusammenhänge wird mit Cramers V beurteilt. Zur Post-hoc-Analyse von Unterschieden werden exakte Fisher-Tests herangezogen (SHAN/GERSTENBERGER 2017). Die Alphafehler-Kumulierung findet über eine Bonferroni-Korrektur Berücksichtigung.

Zur Überprüfung des Zusammenhangs zwischen den RIASEC-Codes und dem Geschlecht fungiert auf globaler Ebene eine MANOVA. Standardfehler werden über parametrischen Bootstrap (n = 5000) ermittelt (KONIETSCHKE u. a. 2015; TODOROV/FILZMOSER 2010). Grund ist die Verletzung der Annahme multivariater Normalverteilung und Varianzhomogenität. Die abhängigen Variablen sind die sechs Ausprägungen der RIASEC-Merkmale, die unabhängige Variable das Geschlecht. Als Post-hoc Tests dienen Mann-Whitney-U-Tests. Auch hier wird eine Bonferroni-Korrektur vorgenommen. Demnach unterscheiden sich die ausgeübten Berufe bei Männern und Frauen hinsichtlich einer RIASEC-Ausprägung, z. B. ‚realistic', signifikant, wenn $p < .05/6$ gilt.

Hypothese 2
Im Fall der Operationalisierung des Niveaus über die intervallskalierte Variable ‚Prestige' kommt eine lineare Regression zum Einsatz, im Fall der KldB-Levels ein ordinales Logit-Modell. Eine wichtige Annahme des ordinalen Logit-Modells ist die von ‚proportional-odds', d. h. die Parameterschätzungen divergieren nicht zwischen den einzelnen Schwellen. Diese Annahme wird überprüft mit einem Brant-Test (BRANT 1990). Sollte die Annahme verletzt sein, wird ein ‚Generalized Ordered Logistic Model' verwendet (WILLIAMS 2006). Hierbei werden für das Überschreiten der einzelnen Schwellen, d. h. für den Übergang von Level 2 auf 3 und von 3 auf 4, jeweils eigene Parameter geschätzt.

Ausreißer können die Validität der Regressionsparameter bei Verwendung der Kleinste-Quadrate-Methode beeinträchtigen (ROUSSEEUW/LEROY 2005, S. 8). Aus diesem Grund wird ein SMDM-Schätzer verwendet (KOLLER/STAHEL 2011; MAECHLER u. a. 2018). Der SMDM-Schätzer ist ein robustes Verfahren, das in Simulationen im Vergleich zu MM-Schätzern überlegene Ergebnisse lieferte (KOLLER/STAHEL 2011, S. 2510 ff.). Die Nutzung robuster Verfahren scheint von besonderer Bedeutung, weil die Korrektheit der Angaben der XING-Nutzer nicht garantiert werden kann. Dieser Punkt wird in den Limitationen nochmals aufgegriffen. Zur Beurteilung der statistischen Signifikanz der Ergebnisse dienen heteroskedastizitäts- (und autokorrelationsrobuste) Standardfehler (CROUX/DHAENE/HOORELBEKE 2003).

Um mögliche Moderationseffekte zu berücksichtigen, werden sukzessive Interaktionsterme in die Regressionsmodelle aufgenommen. Auf diese Weise kann nicht nur geprüft werden, ob z. B. ein Einfluss des Geschlechts auf das Niveau global vorliegt, sondern nur bei Vorliegen bestimmter Abschlüsse.

Hypothese 3
Zur Prüfung der Einflüsse auf die Wahrscheinlichkeit einer Selbstständigkeit dient ein robustes GLM mit Logit-Link (MAECHLER u. a. 2018). Analog zu Hypothese 2 werden sukzessive Interaktionsterme in das Modell aufgenommen.

4 Ergebnisse

4.1 Berufsstruktur: Berufsfachlichkeit

Die in abhängigen Beschäftigungsverhältnissen ausgeübten Berufe finden sich in Tab. 1. Aufgenommen wurden alle Berufsgruppen, die einen Anteil von mindestens 1 % oder gleichbedeutend mehr als 20 Nennungen aufweisen. Mit diesen 15 Berufsgruppen lassen sich 90.22 % aller in der Stichprobe identifizierten Berufe abdecken. Als Vergleichswerte dienen Daten der Bundesagentur für Arbeit. Daraus wird eine relative Häufigkeit gebildet (rHD). Den Nenner bildet die Anzahl aller abhängig Beschäftigten in Deutschland, die in einer Berufsgruppe tätig sind, die von mindestens einer Person in der Stichprobe (n = 2058) besetzt wird. Dadurch werden die für Wirtschaftspädagogen/-innen kaum relevanten Berufsgruppen wie z. B. ‚Reinigung' ausgeschlossen. Der Zähler ist die Anzahl der abhängig Beschäftigten in der betreffenden Berufsgruppe in Deutschland. Durch die ermittelte Kennzahl kann beurteilt werden, ob eine Überrepräsentation von Wirtschaftspädagogen/-innen in bestimmten Berufsgruppen vorliegt.

Tab. 1. Nichtselbstständig ausgeübte Berufe außerhalb des Schuldiensts (n = 2058).

Rang	KldB	Berufsgruppe	rH/rHD	(m/w)[a]
1	715	Personalwesen und -dienstleistung	20.89 %/1.08 %	23.95/76.05 %***
2	713	Unternehmensorganisation und -strategie (inkl. Projektleitung und Unternehmensberatung [Anmerkung Autor])	19.83 %/ 8.10 %	51.47/48.53 %***
3	722	Rechnungswesen, Controlling und Revision	8.41 %/2.28 %	41.04/58.96 %
4	921	Werbung und Marketing	7.53 %/2.20 %	43.87/56.13 %
5	843	Lehr- und Forschungstätigkeit an Hochschulen	6.46 %/1.04 %	50.38/49.62 %
6	611	Einkauf und Vertrieb	4.66 %/3.63 %	62.50/37.50 %***
7	711	Geschäftsführung und Vorstand	4.52 %/1.01 %	64.52/35.48 %***
8	721	Versicherungs- und Finanzdienstleistung	4.47 %/3.94 %	43.48/56.52 %
9	844	Lehrtätigkeit an außerschulischen Bildungseinrichtungen (inkl. Bildungsmanagement und Coaching [Anmerkung Autor])	3.01 %/ 0.38 %	37.10/62.90 %
10	714	Büro- und Sekretariat	2.72 %/10.72 %	16.07/83.93 %**
11	723	Steuerberatung	2.33 %/0.93 %	25.00/75.00 %
12	621	Verkauf (ohne Produktspezialisierung)	1.60 %/ 6.44 %	39.39/60.61 %
13	432	IT-Systemanalyse, IT-Anwendungsberatung und IT-Vertrieb	1.55 %/ 0.84 %	46.88/53.12 %
14	433	IT-Netzwerktechnik, IT-Koordination, IT-Administration und IT-Organisation	1.17 %/0.78 %	79.17/20.83 %**
15	924	Redaktion und Journalismus	1.07 %/ 0.42 %	36.36/63.64 %
Summe			90.22 %/43.79 %	

a Aufteilung in m = männlich und w = weiblich. Frauenanteil insgesamt 58.02 %. Anmerkungen. rH = relative Häufigkeit. rHD = relative Häufigkeit dieser Berufsgruppe bezogen auf die von den 2058 Personen genannten Berufe in Deutschland (BUNDESAGENTUR FÜR ARBEIT 2017). Signifikanzniveau für die Geschlechterunterschiede unter Berücksichtigung einer Bonferroni-Korrektur: *** p < .001/15, ** p < .01/15, * p < .05/15.

Auf globaler Ebene besteht ein signifikanter Zusammenhang zwischen Geschlecht und Berufsfachlichkeit (p < .001). Er ist mittelstark ausgeprägt: Cramers V = .307. Die Post-hoc-Tests ergaben folgende signifikante Unterschiede: Männer sind häufiger beschäftigt in ‚Unternehmensorganisation und -strategie', ‚Einkauf und Vertrieb', ‚Geschäftsführung und Vorstand' sowie ‚IT-Netzwerktechnik, IT-Koordination, IT-Administration und

IT-Organisation'. Frauen arbeiten signifikant häufiger in ‚Personalwesen und -dienstleistung' sowie ‚Büro und Sekretariat'. Die Ergebnisse können als Evidenz für H1a gelten.

4.2 Berufsstruktur: Interessenprofile

In Tab. 2 findet sich die Ausprägungen der RIASEC-Merkmale der ausgeübten Berufe der abhängig Beschäftigten, gruppiert nach Geschlecht.

Tab. 2. Ausprägung der RIASEC-Merkmale der ausgeübten Berufe gruppiert nach Geschlecht (n = 2058).

Mittelwert (Standardabweichung)	R	I	A	S	E	C
Männlich	2.15 (0.94)	3.10 (1.51)	2.14 (0.99)	3.34 (1.48)	5.79 (1.50)	4.82 (1.10)
Weiblich	2.08 (0.94)	3.08 (1.06)	2.13 (1.06)	3.42 (1.58)	5.46 (1.43)	5.02 (1.32)
Gesamt	2.11 (0.94)	3.09 (1.50)	2.14 (1.03)	3.39 (1.54)	5.60 (1.47)	4.93 (1.23)

Anmerkungen. R = ‚realistic', I = ‚investigative', A = ‚artistic', S = ‚social', E = ‚enterprising', C = ‚conventional'. Gesamtmittelwerte über alle Berufe hinweg basierend auf einer Skala von 1 bis 7.

Die Ergebnisse der robusten MANOVA sind: $\chi^2 = 66.597$, $df = 6$, $p < .001$, $\eta^2 = .036$, Wilk`s $\lambda_{robust} = .910$. Insgesamt ist der Zusammenhang zwischen dem Geschlecht und den Ausprägungen der RIASEC-Merkmale der ausgeübten Berufe zwar signifikant aber eher schwach ausgeprägt. Konform mit H1b weisen die von Wirtschaftspädagogen ausgeübten Berufe im Vergleich zu jenen von Wirtschaftspädagoginnen eine höhere Ausprägung bei ‚enterprising' auf ($p < .001$). Die von Frauen angegebenen Berufe zeigen eine signifikant höhere Ausprägung bei ‚conventional' ($p = .004$). Das ist Evidenz für H1b. Konträr zu H1b findet sich allerdings kein Unterschied hinsichtlich ‚realistic' und ‚investigative'. Die Befunde zu H1b sind damit insgesamt gemischt.

4.3 Berufsniveau: Level nach KldB

Wirtschaftspädagogen/-innen, die einen Beruf auf dem ersten der vier Level der KldB ausüben, finden sich nicht in der Stichprobe. 89 Personen sind auf Level 2 (4.33 % [m = 21.35 %, w = 78.65 %]), 538 auf Level 3 (26.14 % [m = 37.35 %, w = 62.65 %]) und 1431 auf Level 4 (69.53 % [m = 45.00 %, w = 55.00 %]) beschäftigt. Der Median des Levels liegt für die fünf Studienabschlüsse grundsätzlich auf Level 4; Ausnahme ist der Bachelor auf Level 3.

Die angemessene Form der ordinalen Regression ist das ‚Generalized Ordered Logistic Model', weil proportionale Chancenverhältnisse (odds-ratios) nicht vorliegen ($\chi^2(df) = 18.79(7)$, $p = .010$). Post-hoc-Tests zeigen als Grund für die Verletzung dieser Eigenschaft die Variable ‚Geschlecht' ($\chi^2(df) = 8.01(1)$, $p < .001$). Die Ergebnisse der Modellschätzung finden sich in Tab. 3. Das Pseudo-R^2 nach Nagelkerke beträgt .084. Interaktionsterme konnten aufgrund von Schätzproblemen nicht aufgenommen werden. Das trifft auch zu, wenn für die zwei Schwellenschätzungen jeweils separate (robuste) GLM mit Logit-Link verwendet werden. Die Robustheit der Ergebnisse in Tab. 3 konnte mit dieser Methode allerdings bestätigt werden.

Tab. 3. ‚Generalized Ordered Logistic Modell' für KldB-Level (n = 2058).

	Logit 2|3 (B1)	e^{B1}	s. e. B1	p > |z|	Logit 3|4 (B2)	e^{B2}	s. e. B2	p > |z|
Intercept	2.409		0.533	.000***	-0.231		0.250	.354
Diplom	1.016	2.762	0.333	.002**	0.806	2.264	0.192	.000***
Magister	1.562	4.768	0.510	.000***	0.765	2.149	0.217	.000***
Master	1.562	4.060	0.319	.000***	0.732	2.079	0.187	.001***
Promotion	NA	NA	NA	NA	2.799	16.423	0.541	.000***
Erfahrung	0.091	1.095	0.057	.115	0.068	1.071	0.027	.012*
Erfahrung²	-0.002	0.998	0.002	.349	-0.000	1.000	0.001	.649
weiblich	-0.971	0.379	0.274	.000***	-0.190	0.827	0.107	.076
Schweiz	-0.411	0.662	0.785	.712	0.110	1.116	0.325	.736
Deutschland	-0.367	0.693	0.382	.331	-0.225	0.798	0.150	.130

Anmerkungen. VIF < 4. Bachelor als Referenzgruppe beim Abschluss, Österreich beim Land des Studienabschlusses. Logit 2|3 Schätzparameter für den Übergang von Stufe 2 auf 3. e = Eulersche Zahl, s. e. = Standardfehler. Keine Person mit Promotion auf Stufe 2, entsprechende Parameter nicht schätzbar. *** p < .001, ** p < .01, * p < .05.

Im Vergleich zum Bachelor weisen Personen mit Diplom, Magister und Master eine signifikant höhere Wahrscheinlichkeit auf, einen Beruf auf Level 3 statt 2 auszuüben ($p < .002$). Beispielsweise ist die Chance im Fall eines Masters im Vergleich zum Bachelor 4.060-mal so hoch. Eine Schätzung für Promovierte ist nicht möglich, weil niemand mit einem solchen Abschluss auf Level 2 tätig ist. Ein Beruf auf dem Level 4 statt 3 ist im Vergleich zum Bachelor für alle anderen Abschlüsse signifikant wahrscheinlicher ($p < .001$). Die Chance ist beispielsweise im Fall des Masters 2.079-mal so hoch. Darüber hinaus ist es mit Promotion im Vergleich zu allen anderen Abschlüssen signifikant wahrscheinlicher, einen Beruf auf Level 4 statt 3 auszuüben ($p < .001$). Die Chance im Vergleich zum Master ist beispielsweise 7.899-mal (16.423/2.079) so hoch. Die Ergebnisse sind allesamt Evidenz für H2a.

Belege für einen unterlinearen Einfluss der Erfahrung auf das Level gibt es nicht. Die quadratischen Terme sind nicht signifikant. Der lineare Term ist nur für den Übergang von Level 3 auf 4 signifikant ($e^{B_2} = 1.071$, $p = .012$). Aufgrund der Mittelwertzentrierung bedeutet das: Ausgehend von einer zehnjährigen Berufserfahrung ist bei einer Erhöhung um eine marginale Einheit die Chance einen Beruf auf Level 4 statt auf Level 3 auszuüben 1.071-mal so hoch. Zusammenfassend sprechen die Ergebnisse eher gegen H2b.

Hinsichtlich des Geschlechts ist die Chance für einen Beruf auf Level 3 statt 2 bei Frauen im Vergleich zu Männern geringer ($e^{B_1} = 0.379$, $p < .001$). Für den Übergang von Level 3 auf 4 ist der Unterschied nicht signifikant ($e^{B_2} = 0.827$, $p = .076$). Die Evidenz für H2c ist damit gemischt. Im unteren Niveaubereich zeigt sich ein signifikant negativer Einfluss des Geschlechts ‚weiblich' auf die Wahrscheinlichkeit des Übergangs, nicht aber im oberen Bereich.

4.4 Berufsniveau: Prestige

Die Ausprägung der Variable ‚Prestige' bei den abhängig Beschäftigten, gruppiert nach Geschlecht und höchstem Abschluss, zeigt Abb. 1.

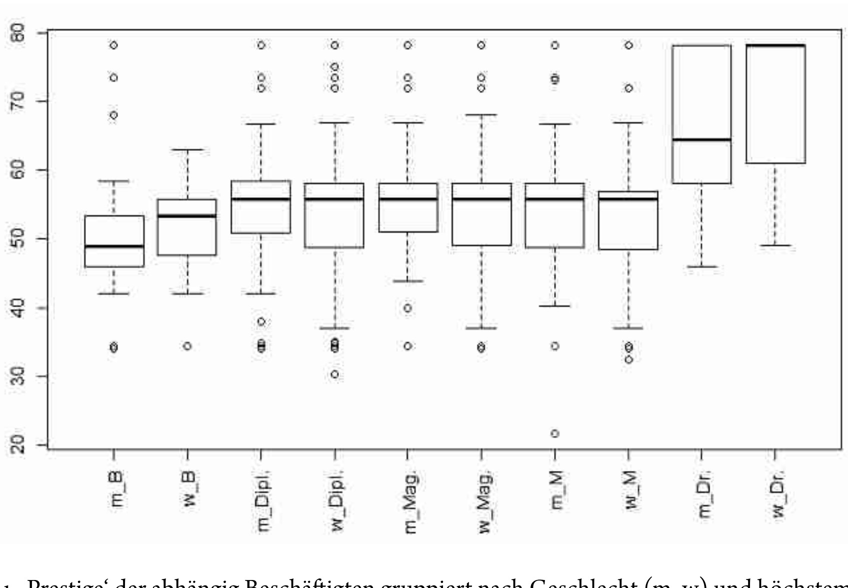


Abb. 1. ‚Prestige' der abhängig Beschäftigten gruppiert nach Geschlecht (m, w) und höchstem Abschluss. B = Bachelor, Dipl. = Diplom, Mag. = Magister, M = Master, Dr. = Promotion. Skalenminimum = 12 (Schuhputzer), Skalenmaximum = 78 (Universitätsprofessor/-in).

Ergebnisse der Regressionsanalyse finden sich in Tab. 4. Aufgrund des Logarithmieren der abhängigen Variable können die Regressionsparameter als prozentuale Veränderungen interpretiert werden.

Tab. 4. Robuste Regression für logarithmiertes ‚Prestige' (n = 2058).

Variable	Modell 1		Modell 2		Modell 3	
	b	s. e.	b	s. e.	b	s. e.
Intercept	3.939	0.013***	3.914	0.018***	3.907	0.018***
Diplom	0.057	0.011***	0.081	0.017***	0.089	0.018***
Magister	0.047	0.013***	0.082	0.019***	0.091	0.020***
Master	0.050	0.011***	0.086	0.019***	0.084	0.019***
Promotion	0.271	0.017***	0.265	0.023***	0.273	0.024***
weiblich (w)	-0.002	0.006	0.038	0.020	0.055	0.022*
Erfahrung	0.003	0.001***	0.003	0.001***	0.001	0.001
$Erfahrung^2$	0.000	0.000	0.000	0.000	0.000	0.000
Schweiz	0.035	0.017*	0.036	0.017*	0.037	0.017*
Deutschland	-0.001	0.008	-0.002	0.008	-0.001	0.008
Diplom*w			-0.038	0.022	-0.055	0.023
Magister*w			-0.057	0.023*	-0.075	0.025**
Master*w			-0.056	0.024*	-0.054	0.024*
Doktor*w			0.036	0.034	0.020	0.035
Erfahrung*w					0.003	0.001*
$Erfahrung^2$*w					0.000	0.000
	R^2 = .151		R^2 = .155		R^2 = .156	

Anmerkung. VIF < 4. Bachelor als Referenzgruppe beim Abschluss, Österreich beim Land des Studienabschlusses. s. e. = Standardfehler. *** p < .001, ** p < .01, * p < .05.

Im Vergleich zum Bachelor beeinflussen alle anderen Abschlüsse das Prestige des ausgeübten Berufs signifikant positiv. Im Fall eines Masterabschlusses ist ein 5.0 % höheres Prestige zu erwarten (p < .001). Im Fall einer Promotion ist mit einem im Vergleich zu allen anderen Abschlüssen höheren Prestige zu rechnen, beispielsweise im Vergleich zum Master um 22.1 % (27,1 %-5.0 %, p < .001). Diese Aussage trifft grundsätzlich auch zu, wenn das Geschlecht kontrolliert wird (Modell 2), d. h. sowohl bei Männern als auch bei Frauen ist bei höherem Abschluss mit einem höheren Prestige zu rechnen. Frauen können bei einem Masterabschluss im Vergleich zum Bachelor ein 3.0 % (8.6 %-5.6 %) höheres Prestige erwarten, Männer 8.6 % (p = .037 bzw. < .001). Weiterhin ist bei Männern mit Promotion im Vergleich zum Master ein um 17.9 % (26.5 %-8.6 %) höheres Prestige zu erwarten, bei Frauen ein um 27.1 % (17.9 %+5.6 %+3.6 %) höheres Prestige (p jeweils < .001). Einzige Ausnahme sind weibliche Magister: Hier ist im Vergleich zu weiblichen Bachelorabsolventen kein signifikant höheres Prestige zu erwarten (p = .100). Allerdings sind Magister und Bachelor nicht direkt vergleichbare Abschlüsse. Insgesamt wird H2a damit angenommen.

Global betrachtet (Modell 1) führt ausgehend von einer Erfahrung von zehn Jahren ein zusätzliches Jahr zu einem ca. 0.3 % höheren erwarteten Prestige. Evidenz für einen nichtlinearen Einfluss gibt es nicht (Modell 1 und 2). Bei Kontrolle des Geschlechts (Modell 3) zeigt sich ein signifikant positiver linearer Term bei Frauen ($p < .001$): Ausgehend von einer Erfahrung von zehn Jahren führt ein weiteres Jahr bei Wirtschaftspädagoginnen zu einer erwarteten Zunahme des Prestiges um ca. 0.4 % (0.1 %+0.3 %). Bei Männern ist der lineare Term hingegen nicht signifikant ($p = .186$). Belege für einen nichtlinearen Zusammenhang zwischen Erfahrung und Prestige gibt es auch nach Kontrolle des Geschlechts nicht. Insgesamt ist die Evidenz bei H2b damit gemischt. Zwar zeigt sich insgesamt ein Einfluss der Erfahrung auf das Prestige, dieser ist allerdings nicht unterlinear.

Global betrachtet (Modell 1) hat das Geschlecht keinen Einfluss auf das Prestige. Bei Kontrolle des Abschlusses (Modell 2) ist allerdings im Fall einer Promotion bei Frauen im Vergleich zu Männern ein um 7.4 % (3.8 %+3.6 %) höheres Prestige zu erwarten ($p = .007$). Das gilt auch im Fall des Bachelors, allerdings nur im Fall von Modell 3. Frauen können hier mit einem 5.5 % höheren Prestige als Männer rechnen ($p = .022$); ohne Kontrolle der Interaktion von Erfahrung und Geschlecht ist der Unterschied nicht signifikant (Modell 2: $p = .052$). Insgesamt treten bei zwei von fünf Abschlüssen Geschlechterunterschiede zu Gunsten von Frauen auf. Die Befunde sprechen gegen H2c, weil ein höheres Prestige bei Männern vermutet wurde.

4.5 Selbstständigkeit

Die häufigsten selbstständigen Tätigkeiten finden sich in Tab. 5. Auf globaler Ebene ergeben sich Geschlechterunterschiede hinsichtlich der Berufsfachlichkeit der selbstständig ausgeübten Tätigkeiten ($p < .001$). Der Zusammenhang ist mittel ausgeprägt, Cramers $V = .331$. In Post-hoc-Tests zeigte sich lediglich ein signifikanter Unterschied: Frauen sind häufiger Trainer/-in als Männer ($p = .002$). Insgesamt wird H3a damit angenommen.

Tab. 5. Selbstständig ausgeübte Tätigkeiten von Wirtschaftspädagogen/-innen (n = 378).

Rang	KldB	Tätigkeit	aH	rH	(m/w)[a]
1	71324	Unternehmensberater/-in	111	29.37 %	63.06/36.94 %
2	71104	Unternehmer/-in	76	20.11 %	64.47/35.53 %
3	84294	Trainer/-in	45	11.90 %	35.56/64.44 %*
4	72124	Coach	35	9.26 %	37.14/62.86 %
5	84304	Dozent/-in	19	5.03 %	47.37/52.63 %
6	84404	Vermögensberater/-in	14	3.70 %	85.71/14.29 %
7	72304	Steuerberater/-in	7	1.85 %	57.14/42.86 %
Summe			307	81.22 %	

a Aufteilung m = männlich und w = weiblich. Verteilung in Stichprobe: Männer = 57.41 %, Frauen 42.59 %. Anmerkungen. aH = absolute Häufigkeit, rH = relative Häufigkeit. * $p < .05/7$ (Bonferroni-Korrektur).

Ersichtlich wird aus Tab. 5, dass alle selbstständig ausgeübten Tätigkeiten auf dem Level 4 der KldB zu verorten sind. In Tab. 6 finden sich die Einflussgrößen auf die Wahl einer selbstständigen Tätigkeit.

Tab. 6. Robustes GLM mit Logit-Link für Selbstständigkeit (n = 2436).

Variable	Modell 1		Modell 2		Modell 3	
	Logit	s. e.	Logit	s. e.	Logit	s. e.
Intercept	-3.312	1.023***	-3.480	1.264**	-2.508	0.819**
Diplom	1.740	1.008	1.843	1.253	1.194	0.805
Magister	1.311	1.024	1.521	1.271	0.838	0.832
Master	2.688	1.034**	3.677	1.289**	2.139	0.844*
Promotion	2.212	1.027*	2.372	1.272	1.642	0.834*
weiblich (w)	0.139	0.142	0.979	1.760	0.392	1.803
Erfahrung	0.191	0.017***	0.194	0.017***	0.116	0.020***
$Erfahrung^2$	-0.004	0.001***	-0.004	0.001***	-0.001	0.001
Schweiz	-0.798	0.434	-1.000	0.436*	-0.856	0.423
Deutschland	-0.467	0.208*	-0.525	0.209*	-0.444	0.213*
Diplom*w			-0.619	1.766	-0.751	1.823*
Magister*w			-0.933	1.776	-0.895	1.830
Master*w			-2.053	1.803	-0.310	1.854
Doktor*w			-0.688	1.821	-0.733	1.879
Erfahrung*w					0.208	0.040***
$Erfahrung^2$*w					-0.009	0.002***
	$R^2 = .128$		$R^2 = .133$		$R^2 = .144$	

Anmerkung. Pseudo-R^2 nach Nagelkerke. Referenzkategorie Bachelor beim Abschluss, Österreich beim Land. s. e. = Standardfehler. *** p < .001, ** p < .01, * p < .05.

Global betrachtet (Modell 1) ist im Vergleich zum Bachelor die Wahrscheinlichkeit einer Selbstständigkeit bei einem Masterabschluss signifikant höher ($e^{Logit} = 14.702$, p = .009). D. h. die Chance von Personen mit Master selbstständig tätig zu sein, ist 14.702-mal so hoch wie für jene mit Bachelorabschluss. Mit einer Promotion wiederum ist im Vergleich zum Bachelor ($e^{Logit} = 9.134$, p = .031), Diplom ($e^{Logit} = 1.603$, p = .046) und Magister ($e^{Logit} = 2.462$, p = .001) eine signifikant höhere Wahrscheinlichkeit einer Selbstständigkeit verbunden. Nicht signifikant ist der Unterschied zum Master (p = .151). Wird das Geschlecht berücksichtigt (Modell 2), zeigt sich nur bei Männern mit Master im Vergleich zum Bachelor eine signifikant höhere Wahrscheinlichkeit einer Selbstständigkeit ($e^{Logit} = 39.528$, p = .004). Insgesamt ist die Evidenz für H3b damit gemischt. Zu beachten ist in diesem Kontext die geringe statistische Power aufgrund der sehr geringen Anzahl an selbstständigen Personen mit Bachelor.

Hinsichtlich der Erfahrung gibt es auf globaler Ebene (Modell 1) Evidenz für einen konkaven Zusammenhang zwischen der Chance einer Selbstständigkeit und der Erfahrung. Der lineare Term ist positiv: Ausgehend von einer Erfahrung von 11.19 Jahren (Mittelwert Selbstständige und Nichtselbstständige) ist bei einem zusätzlichen Jahr die Chance einer Selbstständigkeit 1.210-mal so hoch ($p < .001$). Der quadratische Term ist signifikant negativ ($e^{Logit} = 0.996$, $p < .001$). Es bestehen Geschlechterunterschiede im Einfluss der Erfahrung auf die Wahrscheinlichkeit einer Selbstständigkeit (Modell 3). Der lineare und quadratische Term unterscheidet sich jeweils signifikant bei Männern und Frauen ($p = .040$ bzw. $.002$). Bei Wirtschaftspädagogen beträgt der exponierte Logit des linearen Terms 1.123 ($p < .001$) und des quadratischen Terms 0.999 ($p = .106$). Bei Wirtschaftspädagoginnen beträgt der exponierte Logit des linearen Terms 1.383 ($e^{(0.116+0.208)}$, $p < .001$), des quadratischen Terms 0.992 ($e^{(-0.001-0,009)}$, $p < .001$). D. h. bei Frauen mit 12.19 statt 11.19 Jahren Berufserfahrung ist die Chance einer Selbstständigkeit ca. 1.383-mal so hoch. Insgesamt sprechen die Ergebnisse für einen konkaven Zusammenhang zwischen Erfahrung und Wahrscheinlichkeit einer Selbstständigkeit bei Frauen. Bei Männern gibt es nur Evidenz für einen linearen Zusammenhang. Vergleicht man die Kurvenverläufe im relevanten Bereich einer Berufserfahrung von 0 bis 40 Jahren, zeigt sich ein deutlich stärker positiver Einfluss der Erfahrung auf die Chance einer Selbstständigkeit bei Frauen im Vergleich zu Männern. Die Evidenz für H3c ist gemischt. Zwar zeigt sich durchgehend ein positiver Einfluss der Erfahrung auf die Chance einer Selbstständigkeit, dieser ist allerdings nur bei Frauen konkav.

Evidenz für einen Einfluss des Geschlechts auf die Wahrscheinlichkeit einer Selbstständigkeit liefert keines der drei Modelle. Zwar besteht ein Zusammenhang zwischen Geschlecht und Häufigkeit einer Selbstständigkeit, wenn nur diese beiden Variablen berücksichtigt werden ($p < .001$). Der Zusammenhang verschwindet aber bei Kontrolle der Berufserfahrung. H3d ist damit abzulehnen.

5 Zusammenfassung und Diskussion

5.1 Berufsstruktur

Hypothese 1a wurde angenommen. D. h. es gibt Belege für einen Zusammenhang zwischen dem Geschlecht und der Berufsfachlichkeit (Cramers $V = .307$). Am häufigsten sind Wirtschaftspädagogen/-innen außerhalb des Schuldienstes in der Berufsgruppe ‚Personalwesen und -dienstleistung' abhängig beschäftigt. Dabei handelt es sich zu 76 % um Frauen, bei einem Gesamtfrauenanteil in der Stichprobe von 58 %. Die Anteile sind nicht ungewöhnlich. Die Arbeitsmarktstatistik der Bundesagentur für Arbeit in Deutschland weist – Stand 31.12.2016 – einen Frauenanteil in dieser Berufsgruppe von 72 % aus (Bundesagentur für Arbeit 2017). Betrachtet man die leitenden Tätigkeiten im Personalbereich (Berufsuntergruppe 7519), sinkt der Frauenanteil auf 61 % (Deutschland gesamt 52 %). In eine ähnliche Richtung geht der mit 35 % signifikant niedrigerer Frauenanteil in der Berufsgruppe ‚Geschäftsführung und Vorstand'

(Deutschland gesamt 22 %). Weiter auffällig ist ein Frauenanteil in der Berufsgruppe ‚Büro und Sekretariat' von 84 % (Deutschland 80 %). Aufgrund der Bezeichnung wäre hier nicht unbedingt ein wirtschaftspädagogisches Tätigkeitsfeld zu erwarten. Mit deutlichem Abstand am häufigsten verbirgt sich dahinter eine Berufsangabe wie Assistent/-in der Geschäftsführung, des Vorstands o. Ä. Dabei handelt es sich um durchaus anspruchsvolle Tätigkeiten (Level 3 nach KldB).

Bei der Analyse der Berufsfachlichkeit stellt sich auch die Frage, ob die im Basiscurriculum genannten wirtschaftspädagogischen Tätigkeitsfelder mit den tatsächlich besetzten korrespondieren. Der Abgleich wird dadurch erschwert, dass das Basiscurriculum nicht der KldB oder einer anderen Berufsklassifikation folgt. Was sich zeigt ist, dass die Berufsgruppen ‚Personalwesen und -dienstleistung' (20.89 % vs. 1.08 %), ‚Lehr- und Forschungstätigkeit an Hochschulen' (6.46 % vs. 1.04 %) und ‚Lehrtätigkeit an außerschulischen Bildungseinrichtungen' (3.01 % vs. 0.38 %) im Vergleich zur Situation am deutschen Arbeitsmarkt bei Wirtschaftspädagogen/-innen überrepräsentiert sind. Diese im Basiscurriculum explizit genannten Tätigkeitsfelder sind damit in der Tat relevante Beschäftigungsfelder. Auffällig ist der mit 4.00 % durchaus beachtliche Anteil von Unternehmensberatern/-innen. Dieses Tätigkeitsfeld wird im Basiscurriculum nicht erwähnt.

Bezogen auf H1b sind die Ergebnisse gemischt. Wie vermutet, ist das Interessenprofil ‚enterprising' bei den von Wirtschaftspädagogen ausgeübten Berufen stärker ausgeprägt, bei Wirtschaftspädagoginnen ‚conventional'. Bei ‚realistic' und ‚investigative' gibt es, anders als erwartet, keine Unterschiede. Insgesamt ist der Zusammenhang zwischen den RIASEC-Merkmalen der ausgeübten Berufe und dem Geschlecht eher schwach ausgeprägt ($\eta^2 = .036$). Hinzu kommt, dass der Holland-Code der ausgeübten Berufe, der sich aus den drei am stärksten ausgeprägten RIASEC-Merkmalen zusammensetzt, bei beiden Geschlechtern ‚ECS' ist. Nach BERGMANN und EDER (2005, S. 122) weisen Wirtschaftspädagogen/-innen auch ‚ECS' als Interesse auf. Somit scheint, zumindest global betrachtet, eine gute Passung zwischen persönlichem Interesse und den Anforderungen der ausgeübten Berufe zu bestehen. Neben Wirtschaftspädagogen/-innen weisen auch Betriebs- und Volkswirte ‚ECS' als Holland-Code auf (PFUHL 2010, S. 52). Das spricht für ähnliche Interessensstrukturen von nicht im Schuldienst tätigen Wirtschaftspädagogen/-innen und weiteren Wirtschaftswissenschaftlern. Mit der Interessensstruktur von Lehrpersonen allgemein ‚SAE' (ROTHLAND 2014, S. 352) besteht eine eher geringe Interessenskongruenz: Der Zener-Schnuelle-Index (mögliche Ausprägung von 0 bis 6) beträgt 2.

5.2 Berufsniveau

Beim Niveau, operationalisiert über die vier KldB-Level, fällt die mit 79 % ausgeprägte Überrepräsentation von Frauen auf Level 2 auf. Zu erwarten wären 58 %. Dieser Befund schlägt sich auch nieder in einer für Frauen im Vergleich zu Männern 62 % geringeren Chance einer Beschäftigung auf Level 3 statt 2. Da Level 2 bei Wirtschaftspädagogen/

-innen eher nicht zu erwarten wäre, wurden die Berufsangaben inspiziert. Mit deutlichem Abstand am häufigsten verbergen sich hinter Level 2 Angaben wie ‚Sachbearbeiter/-in' oder ‚Verwalter/-in'. Hierunter könnten sich auch anspruchsvolle Tätigkeiten verbergen. Die Geschlechterunterschiede lassen sich damit allerdings nicht erklären. Da es sich bei den Berufsangaben um Selbsteinschätzungen handelt, könnte der Grund eine von Frauen im Vergleich zu Männern zurückhaltende Selbsteinschätzung sein. Hierfür gibt es durchaus Evidenz (LUNDEBERG/FOX/PUNĆCOHAŔ 1994; CORRELL 2001). Frauen sind möglicherweise aber tatsächlich auf geringerem Niveau tätig. Dafür spricht, dass die Erziehung von Kindern bei Akademikerinnen ein zentrales Karrierehindernis darstellen kann (ABELE 2004, S. 166 f.). Um dieser Vermutung nachzugehen, wurde der aktuelle und die beiden vorhergehenden Berufe der betreffenden Personen inspiziert. Hier zeigte sich ein auffälliges Muster: Frauen, die gegenwärtig auf Level 2 tätig sind, waren oftmals bereits in Berufen auf höherem Level tätig. Ein Beispiel für eine Reihenfolge ist: 1) „Assistent Geschäftsführung", 2) „Coach, Dozentin" und 3) „Sachbearbeiter Personal". Solche Verläufe könnten für einen Verzicht auf Karriere zugunsten anderer Lebensbereiche, beispielsweise Familie, sprechen. Zusammenfassend gilt es aber festzuhalten, dass es sich bei Wirtschaftspädagogen/-innen auf Level 2 um ein zahlenmäßig eher wenig bedeutsames Phänomen handelt (4 %). Darüber hinaus scheinen die berichteten Ergebnisse auch nicht spezifisch für Wirtschaftspädagogen/-innen.

Bezüglich des Einflusses der Höhe des Studienabschlusses auf das Niveau konnte bis auf eine Ausnahme (weibliche Magister im Vergleich zu weiblichen Bachelorabsolventen kein signifikant höheres Prestige) stets ein deutlich positiver Effekt festgestellt werden. Besonders ausgeprägt ist der Zuwachs im Fall einer Promotion, s. Abb. 1. Personen mit Promotion können im Vergleich zum Masterabschluss mit einem 22 % höheren Prestige rechnen. Insgesamt wurde Hypothese 2a angenommen, weil Magister und Bachelor nicht direkt vergleichbare Studiengänge sind.

Interessant ist das zahlenmäßige Verhältnis von Personen mit Master und Bachelorabschluss (außerhalb des Schuldienstes). Die Befunde sprechen eher dagegen, dass es sich beim Bachelor um den Regelabschluss handelt. Auf eine Person mit Bachelorabschluss kommen 2.56 Personen mit Master. Damit unterscheiden sich Wirtschaftspädagogen/-innen allerdings nicht strukturell von anderen Studierendengruppen.

Hinsichtlich der Erfahrung zeigte sich keine Evidenz für einen unterlinearen Einfluss auf das Niveau bei abhängig Beschäftigten. Der Grund könnte in der insgesamt eher gering ausgeprägten Berufserfahrung liegen. Sie beträgt im Mittel 10 Jahre (Median 9 Jahre). In diesen Bereichen eher geringer Berufserfahrung, ist ein linearer Zusammenhang mit dem Niveau durchaus plausibel. Nach Kontrolle des Geschlechts zeigt sich nur bei Frauen ein positiver Zusammenhang zwischen Erfahrung und Niveau. Der fehlende Einfluss der Erfahrung auf das Niveau bei Männern ist konträr zu H2b. Möglicherweise lassen sich Wirtschaftspädagogen von zu geringer Erfahrung nicht abhalten, wenn sie einen gewünschten Beruf anstreben.

Hypothese 2c, wonach Wirtschaftspädagoginnen auf geringerem Niveau tätig sind, konnte bei Operationalisierung über das Prestige in keinem Punkt bestätigt werden. Vielmehr sind promovierte Wirtschaftspädagoginnen und solche mit Bachelorabschluss auf höherem Niveau tätig als die männliche Vergleichsgruppe. Besonders ausgeprägt ist der Unterschied bei den Promovierten. In der daraufhin durchgeführten Analyse der Berufsangaben fällt auf, dass promovierte Wirtschaftspädagoginnen zu 59.4 % in der Wissenschaft tätig sind, Wirtschaftspädagogen nur zu 32.1 %. Tätigkeiten in der Wissenschaft sind mit sehr hohem Prestige verbunden.

5.3 Selbstständigkeit

Von den nicht im Schuldienst tätigen Wirtschaftspädagogen/-innen sind 16 % selbstständig tätig. Dabei zeigt sich ein mittelstark ausgeprägter Zusammenhang zwischen Tätigkeiten und Geschlecht (Cramers V = .331). H3a wurde somit angenommen. Im Abgleich mit dem Basiscurriculum verwundert bei dieser Gruppe insbesondere die mit 29 % relativ häufige Tätigkeit als Unternehmensberater/-in. Eine derartige Auffälligkeit ist auch bei den abhängig Beschäftigten erkennbar. Möglicherweise eignet sich ein Wirtschaftspädagogikstudium gut, um die für Unternehmensberatung nötigen Kompetenzen zu erwerben.

Hinsichtlich des Zusammenhangs zwischen Wahrscheinlichkeit einer Selbstständigkeit und der Höhe des Studienabschlusses (H3b) sind die Ergebnisse gemischt. Das lässt sich vermutlich auch auf die geringe statistische Power zurückführen. Als auffällige Gruppen zeigen sich Personen mit Bachelorabschluss. Hier sind nur sechs Personen selbstständig tätig, was angesichts der Bedeutung der Erfahrung aber nicht verwundert.

Bei beiden Geschlechtern auffällig ist der stark positive Einfluss der Berufserfahrung auf die Wahrscheinlichkeit einer Selbstständigkeit. Das ist auch vor dem Hintergrund der genannten Tätigkeitsfelder durchaus plausibel. Beispielsweise benötigt eine selbstständige Tätigkeit als Unternehmensberater/-in ein hohes Maß an Erfahrung. Insgesamt wurde H3c angenommen.

Evidenz für einen negativen Einfluss des Geschlechts ‚weiblich' auf die Wahrscheinlichkeit einer Selbstständigkeit konnte nicht gefunden werden. H3d ist damit abzulehnen.

6 Limitationen, Implikationen, Ausblick

6.1 Limitationen

Limitation der Nutzung von XING-Daten ist die Selbstselektion in die Stichprobe. Wirtschaftspädagogen/-innen der Generation ‚Babyboomer' und ‚X' scheinen im Vergleich zur ‚Generation Y' seltener ein XING-Profil zu betreiben. Dafür spricht die (linkssteile) Verteilung der Variable ‚Erfahrung' (Mittelwert = 10, Median = 9). Das ist kritisch zu sehen, weil sich die Generationen hinsichtlich der Karriereaffinität unterscheiden könnten (AKSOY u. a. 2013). Allgemein ist es problematisch, wenn sich

Personen mit XING-Profil systematisch von jenen ohne XING-Profil unterscheiden. Plausibel ist grundsätzlich eine höhere Karriereaffinität von Personen mit XING-Profil im Vergleich zu Personen ohne Profil. Eine Ausnahme wäre, wenn die Person nur ein LinkedIn-Profil betreibt. Hier wäre dann ein im Vergleich zu XING-Nutzern stärkerer Fokus auf den internationalen Markt wahrscheinlich. Auch könnte XING in Branchen und Berufsgruppen unterschiedlich stark verbreitet sein. Einsichtig scheint beispielsweise eine hohe XING-Nutzung von Personen, die im Bereich ‚Personal' beschäftigt sind. Außerdem sind nur die öffentlichen Profile Teil der Stichprobe. Personen, die ein nicht-öffentliches Profil betreiben, haben möglicherweise andere Persönlichkeitseigenschaften, die sich in der Berufswahl niederschlagen könnten.

Obwohl die Nutzungsbedingungen die Profil-Betreiber dazu verpflichten, ist die Korrektheit der Daten nicht sichergestellt. Bewusste Falschangaben sind eher unwahrscheinlich, weil durch die Öffentlichkeit der Profile eine soziale Kontrolle besteht. Wahrscheinlicher sind fehlende Aktualisierungen des Profils. Diese sind besonders im Rahmen der Auswertungen des Berufsniveaus problematisch. Vor diesem Hintergrund kamen stets (ausreißer-)robuste statistische Verfahren zum Einsatz. Problematisch sind auch unterschiedlich realistische Selbsteinschätzungen der eigenen Position zwischen Männern und Frauen, für die es in der Literatur Belege gibt.

Neben der Limitation der Stichprobe bestehen Limitationen bei den Operationalisierungen der Variablen. Das Prestige wurde im internationalen Kontext ermittelt und spiegelt nicht notwendigerweise die aktuelle Situation im deutschsprachigen Raum wider. Allerdings kann durchaus eine einheitliche Einschätzung des Prestiges über verschiedene Kulturen und auch über die Zeit hinweg festgestellt werden (RATSCHINSKI 2009, S. 88). Die Variable ‚Erfahrung' berücksichtigt lediglich die Zeit seit dem höchsten Studienabschluss. Unterbrechungen, beispielsweise für Kinderbetreuung oder durch Arbeitslosigkeit können nicht berücksichtigt werden. Ferner fließt Erfahrung vor dem Studium oder zwischen Studiengängen nicht mit ein. Darüber hinaus wäre die Berücksichtigung weiterer Kontrollvariablen wie Teilzeitbeschäftigung wünschenswert, um eine höhere Varianzaufklärung zu erreichen.

6.2 Implikationen und Ausblick

Bei der Entwicklung von Studiengängen scheint es wichtig, empirische Erkenntnisse einzubeziehen (SÖLL 2017, S. 29 f.). Der vorliegende Artikel kann einen Beitrag dazu leisten, die in Studiengangsbeschreibungen genannten wirtschaftspädagogischen Tätigkeitsfelder auf ein empirisches Fundament zu stellen. Selbiges gilt für die Konkretisierung „praxisnaher Fragestellungen", wie sie im Basiscurriculum angesprochen werden.

Inhaltlich sind die Ergebnisse dazu geeignet, die Breite wirtschaftspädagogischer Tätigkeitsfelder außerhalb des Schuldienstes aufzuzeigen. Die Ergebnisse können dazu dienen, die Polyvalenz und damit die Attraktivität des Studiengangs zu belegen. Tatsächlich bietet sich mit einem Wirtschaftspädagogikstudium eine große Bandbreite an beruflichen Perspektiven. Die besonders stark ausgeprägten Niveauzuwächse im Fall einer Promotion könnten herangezogen werden, um bei Absolventen für eine Tätigkeit in der Wissenschaft und das Anstreben einer Promotion zu werben.

Methodisch zeigt der Beitrag auf, wie es im Kontext der Berufs- und Wirtschaftspädagogik möglich ist, öffentlich verfügbare und zugleich sehr große Datenmengen für Forschungszwecke nutzbar zu machen. Hierbei wurde dargelegt, welche Schritte voll- oder teilautomatisch vorgenommen werden können. Besonders zentral ist in diesem Rahmen die Idee, verschiedene Datenbanken zu verknüpfen. Durch eine initiale Kodierung der Berufe mit der KldB lassen sich den Berufen eine Vielzahl wichtiger Variablen automatisiert zuordnen. Insbesondere die Verknüpfung mit dem O*Net ermöglicht eine sehr umfassende Charakterisierung von Berufen, die für Forschungszwecke nutzbar gemacht werden kann.

Der vorliegende Beitrag untersucht Personen außerhalb des Schuldiensts. Interessant wäre, wie sich diese Gruppen von jener im Schuldienst unterscheidet. Für einen Zusammenhang zwischen Karriereaffinität und dem Anstreben einer Tätigkeit im Schuldienst gibt es durchaus Evidenz (ROTHLAND/TERHART 2010, S. 793 ff.). Gerade für familienorientierte Personen könnten die Schulferien besonders attraktiv sein. Diese beiden Gruppen systematisch zu vergleichen, wäre ein interessanter Ansatz. Auch ein Vergleich von Wirtschaftspädagogen/-innen mit Betriebswirten/-innen wäre vielversprechend.

Literatur

ABELE, A. E. (2004): Beruf – kein Problem, Karriere – schon schwieriger: Berufslaufbahnen von Akademikerinnen und Akademikern im Vergleich. In: ABELE, A. E. / HOFF, E. H. / HOHNER, H. U. (Hrsg.): Frauen und Männer in akademischen Professionen: Berufsverläufe und Berufserfolg. Heidelberg: Asanger, S. 157–182.

ACHATZ, J. (2005): Geschlechtersegregation im Arbeitsmarkt. In: ABRAHAM, M. / HINZ, T. (Hrsg.): Arbeitsmarktsoziologie: Probleme, Theorien, empirische Befunde. Wiesbaden: VS Verlag, S. 263–301.

AKSOY, L. / BOLTON, R. N. / PARASURAMAN, A. / HOEFNAGELS, A. / MIGCHELS, N. / KABADAYI, S. / GRUBER, T. / KOMAROVA LOUREIRO, Y. / SOLNET, D. (2013): Understanding Generation Y and their use of social media. A review and research agenda. In: Journal of Service Management 24, H. 3, S. 245–267.

BÄRMANN, F. ([2]2017): Xing. Erfolgreich netzwerken im Beruf. Frechen: mitp.

BERGMANN, C. / EDER, F. (2005): Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (UST-R) – Revision. Göttingen: Hogrefe.

BRANT, R. (1990): Assessing proportionality in the proportional odds model for ordinal logistic regression. In: Biometrics 46, H. 4, S. 1171–1178.

BRÜCKER, H. / KLINGER, S. / MÖLLER, J. / WALWEI, U. (2012): Handbuch Arbeitsmarkt. Analysen, Daten, Fakten. Bielefeld: Bertelsmann.

BUNDESAGENTUR FÜR ARBEIT (2006): Tätigkeitsbeschreibung von Wirtschaftspädagoge/Wirtschaftspädagogin (Uni). 22.09.2018.

BUNDESAGENTUR FÜR ARBEIT (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/ Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB-2010/Generische-Publikationen/KldB2010-Printversion-Band1.pdf. 22.09.2018.

BUNDESAGENTUR FÜR ARBEIT (2013): Tabellarische Umsteigeschlüssel zur KldB 2010. https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/ Klassifikation-der-Berufe/KldB2010/Arbeitshilfen/Umsteigeschluessel/ Umsteigeschluessel-Nav.html. 11.05.2018.

BUNDESAGENTUR FÜR ARBEIT (2017): Beschäftigte nach Berufen (KldB 2010) (Quartalszahlen). Stichtag: 31. Dezember 2016. https://statistik.arbeitsagentur.de/nn_217688/Statischer-Content/Rubriken/Beschaeftigung/Beschaeftigte/Beschaeftigte-nach-Berufen-Klassifizierung-der-Berufe-2010.html. 22.09.2018.

BUREAU OF LABOR STATISTICS (2012): ISCO SOC Crosswalk. https://www.bls.gov/soc/ISCO_SOC_Crosswalk.xls. 18.09.2018.

CHARLES, M. / BRADLEY, K. (2009): Indulging our gendered selves? Sex segregation by field of study in 44 countries. In: American Journal of Sociology 114, H. 4, S. 924–976.

CHRISTOPH, B. / LEBER, U. / STÜBER, H. (2017): Einkommen von Bachelor- und anderen Hochschulabsolventen. Höhere Abschlüsse zahlen sich mit dem Alter zunehmend aus. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.

CORRELL, S. J. (2001): Gender and the career choice process. The role of biased self assessments. In: American Journal of Sociology 106, H. 6, S. 1691–1730.

CROUX, C. / DHAENE, G. / HOORELBEKE, D. (2003): Robust standard errors for robust estimators. Leuven: K. U. Leuven.

GANZEBOOM, H. B. G. / TREIMAN, D. J. (1996): Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. In: Social Science Research 25, H. 3, S. 201–239.

GANZEBOOM, H. B. G. / TREIMAN, D. J. (2012): International stratification and mobility file: Conversion tools. http://www.harryganzeboom.nl/isco08/index.htm. 22.09.2018.

GATI, I. / ASHER, I. (2001): Prescreening, In-Depth Exploration, and Choice: From Decision Theory to Career Counseling Practice. In: The Career Development Quarterly 50, H. 2, S. 140–157.

GOTTFREDSON, L. S. (1981): Circumscription and compromise. A developmental theory of occupational aspirations. In: Journal of Counseling Psychology 28, H. 6, S. 545–579.

GOTTFREDSON, L. S. (42002): Gottfredson's Theory of Circumscription, Compromise, and Self-Creation. In: BROWN, D. (Hrsg.): Career choice and development. San Francisco: Jossey-Bass, S. 85–148.

GRAMLINGER, F. / SCHLÖGL, P. / STOCK, M. (2008): Berufs- und Wirtschaftspädagogik in Österreich. Oder: Wer „macht“ die berufliche Bildung in AT? In: bwp@, Spezial 3, S. 1–44.

HACHMEISTER, C.-D. / HARDE, M. E. / LANGER, M. F. (2007): Einflussfaktoren der Studienentscheidung. Eine empirische Studie von CHE und EINSTIEG. Gütersloh: CHE Centrum für Hochschulentwicklung.

HAUER, E. / STOCK, M. (2007): Qualifizierung für Wirtschaft oder Schule? – Polyvalenz des Studiums der Wirtschaftspädagogik an den Standorten Wien und Graz. In: bwp@, H. 12, S. 1–17.

HOLLAND, J. L. (31997): Making vocational choices. A theory of vocational personalities and work environments. Lutz: Psychological Assessment Resources.

INSTITUT FÜR BILDUNGSFORSCHUNG DER WIRTSCHAFT (o. J.): Wirtschaftspädagoge/Wirtschaftspädagogin. https://bic.at/berufsinformation.php?beruf&wirtschaftspaedagoge-wirtschaftspaedagogin&brfid=1254. 22.09.2018.

IWP (2018): Ausbildung in Wirtschaftspädagogik der Universität St. Gallen. http://www.unisg.ch/de/studium/master/zusatzabschluesse/zusatzausbildungwirtschaftspaedagogik/curriculum. 22.09.2018.

KOELLINGER, P. / MINNITI, M. / SCHADE, C. (2013): Gender Differences in Entrepreneurial Propensity. In: Oxford Bulletin of Economics and Statistics 75, H. 2, S. 213–234.

KOLLER, M. / STAHEL, W. A. (2011): Sharpening Wald-type inference in robust regression for small samples. In: Computational Statistics & Data Analysis 55, H. 8, S. 2504–2515.

KONIETSCHKE, F. / BATHKE, A. C. / HARRAR, S. W. / PAULY, M. (2015): Parametric and nonparametric bootstrap methods for general MANOVA. In: Journal of Multivariate Analysis 140, S. 291–301.

KULTUSMINISTERKONFERENZ (2017): Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. https://www.kmk.org/fileadmin/Dateien/ veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf. 22.09.2018.

LANDIS, J. R. / KOCH, G. G. (1977): The measurement of observer agreement for categorical data. In: Biometrics 33, H. 1, S. 159–174.

LUNDEBERG, M. A. / FOX, P. W. / PUNĆCOHAŔ, J. (1994): Highly confident but wrong: Gender differences and similarities in confidence judgments. In: Journal of educational psychology 86, H. 1, S. 114–121.

MAECHLER, M. / ROUSSEEUW, P. / CROUX, C. / TODOROV, V. / RUCKSTUHL, A. / SALIBIAN-BARRERA, M. / VERBEKE, T. / KOLLER, M. / CONCEICAO, E. / DI PALMA, M. (2018): robustbase: basic robust statistics. https://cran.r-project.org/web/packages/robustbase/robustbase.pdf. 22.09.2018.

MAI, C.-M. / MARDER-PUCH, K. (2013): Selbstständigkeit in Deutschland. Wiesbaden: Statistisches Bundesamt.

MARANTO, C. L. / RODGERS, R. C. (1984): Does work experience increase productivity? A test of the on-the-job training hypothesis. In: The Journal of Human Resources 19, H. 3, S. 341.

MINCER, J. (1974): Schooling, experience, and earnings. New York: Columbia University Press.

MORTIMER, J. T. / LORENCE, J. (1979): Occupational experience and the self-concept: A longitudinal study. In: Social Psychology Quarterly 42, H. 4, S. 307.

O*NET (2017): O*NET 22.0 Database. https://www.onetcenter.org/database.html. 11.05.2018.

PAULUS, W. / MATTHES, B. (2013): The German Classification of Occupations 2010 – Structure, Coding and Conversion Table. http://doku.iab.de/fdz/reporte/2013/MR_08–13_EN.pdf. 22.09.2018.

PETERSON, N. G. / MUMFORD, M. D. / BORMAN, W. C. / JEANNERET, P. R. / FLEISHMAN, E. A. / LEVIN, K. Y. / CAMPION, M. A. / MAYFIELD, M. S. / MORGESON, F. P. / PEARLMAN, K. / GOWING, M. K. / LANCASTER, A. R. / SILVER, M. B. / DYE, D. M. (2001): Understanding work using the Occupational Information Network (O*NET): Implications for practice and research. In: Personnnel Psychology 54, H. 2, S. 451–492.

PFUHL, N. (2010): Untersuchung zur Bestimmung von typischen Merkmalen des Image von Studienfächern. Münster: Waxmann.

RATSCHINSKI, G. (2009): Selbstkonzept und Berufswahl. Eine Überprüfung der Berufswahltheorie von Gottfredson an Sekundarschülern. Münster: Waxmann.

ROTHLAND, M. (2014): Warum entscheiden sich Studierende für den Lehrerberuf? In: TERHART, E. / BENNEWITZ, H. / ROTHLAND, M. (Hrsg.): Handbuch der Forschung zum Lehrerberuf. Münster: Waxmann, S. 349–385.

ROTHLAND, M. / TERHART, E. (2010): Forschung zum Lehrerberuf. In: TIPPELT, R. / SCHMIDT, B. (Hrsg.): Handbuch Bildungsforschung. Wiesbaden: VS Verlag, S. 791–810.

ROUSSEEUW, P. J. / LEROY, A. M. (2005): Robust regression and outlier detection. Hoboken, NJ: Wiley-Interscience.

SCHWÄR, H. (2018): LinkedIn oder Xing: Studie zeigt, welches Netzwerk sich bei der Bewerbung mehr bezahlt macht. https://www.businessinsider. de/linkedin-oder-xing-studie-zeigt-welches-netzwerk-sich-bei-der-bewerbung-mehr-bezahlt-macht-2018–3?IR=T. 11.05.2018.

SEKTION BWP DER DGFE (2014): Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik im Rahmen berufs- und wirtschaftspädagogischer Studiengänge. http://www.dgfe.de/fileadmin/ OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf. 22.09.2018.

SHAN, G. / GERSTENBERGER, S. (2017): Fisher's exact approach for post hoc analysis of a chi-squared test. In: PloS ONE 12, H. 12, 1–12.
SHAVELSON, R. J. / HUBNER, J. J. / STANTON, G. C. (1976): Self-concept: Validation of construct interpretations. In: Review of Educational Research 46, H. 3, S. 407–441.
SLOANE, P. F. E. / TWARDY, M. / BUSCHFELD, D. ([2]2004): Einführung in die Wirtschaftspädagogik. Paderborn: Eusl-Verl.
SÖLL, M. (2017): Orientierungspotenziale des Basiscurriculums der Berufs- und Wirtschaftspädagogik für die disziplinäre Binnen- und Außenlegitimität. In: SEIFRIED, J. / SEEBER, S. / ZIEGLER, B. (Hrsg.): Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2017. Opladen: Barbara Budrich, S. 29–42.
SPENCE, M. (1973): Job market signaling. In: The Quarterly Journal of Economics 87, H. 3, S. 355–374.
STOCK, M. / FERNANDEZ, K. / SCHELCH, E. (2007): Quo vadis Wirtschaftspädagogik? In: wissenplus 07/08, H. 5, S. 23–28.
TODOROV, V. / FILZMOSER, P. (2010): Robust statistic for the one-way MANOVA. In: Computational Statistics & Data Analysis 54, H. 1, S. 37–48.
WEBER, S. / GUGGEMOS, J. (2018): Berufswahlbereitschaft und -fähigkeit von Migranten und Geflüchteten. In: Unterrichtswissenschaft 46, H. 1, S. 21–42.
WILLIAMS, R. A. (2006): Generalized ordered logit/partial proportional odds models for ordinal dependent variables. In: The Stata Journal 6, H. 1, S. 58–82.
XING AG (2017): Geschäftsbericht 2016. https://corporate.xing.com/de/investor-relations/publikationen/. 14.09.2017.
ZEHETNER, E. / STOCK, M. / SLEPCEVIC-ZACH, P. (2016): Wipäd – und dann? Ergebnisse aus der aktuellen Abso-Befragung am Standort Graz. In: bwp@ 16/17, H. 1, S. 30–33.
ZLATKIN-TROITSCHANSKAIA, O. / BREUER, K. (2010): Wirtschaftspädagogisches Studium an der Uni Mainz. Ein polyvalenter Bachelor und Master of Science in Wirtschaftspädagogik. In: Erziehungswissenschaft 21, H. 40, S. 125–133.
ZÜLL, C. (2016): The Coding of Occupations. Mannheim: GESIS – Leibniz Institute for the Social Sciences.

DR. JOSEF GUGGEMOS, MBR
Institut für Wirtschaftspädagogik (IWP-Digital-Betrieb), Universität St. Gallen,
Guisanstrasse 1a, CH-9010 St. Gallen, josef.guggemos@unisg.ch

Studie 9:

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 106552.

Computers in Human Behavior 115 (2021) 106552

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends

Sabine Seufert [a], Josef Guggemos [a,*], Michael Sailer [b]

[a] School of Management, Institute of Business Education and Educational Management, University of St.Gallen, St. Jakob-Strasse 21, 9000 St. Gallen, Switzerland
[b] Chair of Education and Educational Psychology, Ludwig Maximilian University of Munich, Leopoldstr. 13, 80802 Munich, Germany

ARTICLE INFO

Keywords:
TPACK
Technology
Knowledge
Skills
Attitudes
Professional development for teachers
Artificial intelligence
Augmentation strategies

ABSTRACT

This is the introductory article for the special issue "Technology-related knowledge, skills, and attitudes of pre- and in-service teachers". It (1) specifies the concept of technology-related knowledge, skills, and attitudes (KSA) of teachers, (2) presents how these KSA are currently assessed, and (3) outlines ways of fostering them among pre- and in-service teachers. The eight articles in the special issue are structured accordingly, and we demonstrate how they contribute to knowledge in these three areas. Moreover, we show how the afterword to the special issue widens the perspective on technology integration by taking into account systems and cultures of practice. Due to their quantitative empirical nature, the eight articles investigate technology at the current state of the art. However, the potential of artificial intelligence has not yet been fully exploited in education. We provide an outlook on potential developments and their implications on teachers' technology-related KSA. To this end, we introduce the concept of augmentation strategies.

1. Introduction

"Emergency remote teaching" has become a worldwide phenomenon due to COVID-19, resulting in a temporary shift to online teaching (Hodges, Moore, Lockee, Trust, & Bond, 2020). These exceptional circumstances have brought the use of technology in education to the attention of a broader public. The term (educational) technology frequently refers implicitly to digital technology. For instance, Tamim, Bernard, Borokhovski, Abrami, and Schmid (2011) include in their second-order meta-analysis about technology "computer technology as a supplement for in-class instruction" (p. 7). The recent meta-analysis of J-PAL (2019), which aims at a comprehensive overview of educational technology, classified four groups: 1) access to technology, e.g., computers and internet access, 2) computer assisted learning, e.g., educational software, 3) technology-enabled behavioral interventions, e.g., gamification, and 4) online learning, e.g., massive open online courses. Regular claims are that (advanced) educational technology will improve learning efficiency, facilitate greater focus on the future professional needs of learners, and foster personality development in a digital society. However, such claims are often based on 'myths' instead of sound research (Kirschner & van Merrienboer, 2013). The meta-analysis of J-PAL (2019)[1] based on evidence from experimental research indicates that the use of technology can lead to positive but not overwhelming effects on learning outcomes. This is in line with the study by Tamim et al. (2011) that condensed forty years of research about technology in education. It reveals average effect sizes between 0.30 and 0.35 (depending on the method used). Both studies reveal a substantial variance among effect sizes, which may indicate that it is not meaningful to talk about technology in general terms. In this vein, the fact that technology use in empirical studies is often of short duration should also be considered. It is questionable whether conditions developed for an experimental design of short duration can be sustained in normal classroom settings (Cheung & Slavin, 2013). Moreover, technology is a very dynamic concept (Koehler, Mishra, & Cain, 2013) and more recent technology may yield higher learning outcomes. In summary, there seems to be some potential for the use of technology in terms of improving educational processes. However, the effectiveness and efficiency of such technology (not surprisingly) depends on the way it is used, which could point to an important role for teachers.

Indeed, the available evidence may indicate an important role of

* Corresponding author.
E-mail address: josef.guggemos@unisg.ch (J. Guggemos).

[1] The Abdul Latif Jameel Poverty Action Lab (J-PAL) is a global research center working to reduce poverty by ensuring that policy is informed by scientific evidence.

https://doi.org/10.1016/j.chb.2020.106552
Received 31 May 2020; Received in revised form 16 August 2020; Accepted 4 September 2020
Available online 6 September 2020

teachers in the integration of technology into education (Fraillon, Ainley, Schulz, Friedman, & Duckworth, 2019; OECD, 2015). In this vein, Dillenbourg (2013) coined the term *orchestration*. Teachers have to manage multiple activities while considering multiple constraints. They have to decide in which context they use or do not use a specific technology. Such decision-making processes may be guided by findings about high-quality learning environments (e.g., Merrill, 2002). As Sweller (2020, p. 1) posited: "Technology-based instruction used without reference to the instructional design principles that flow from human cognition is likely to be random in its effectiveness." To deliberately use technology, teachers need specific knowledge, skills, and attitudes (KSA) (Kirschner, 2015).

In this introductory article for the special issue "Technology-related knowledge, skills, and attitudes of pre- and in-service teachers", we specify the concept of technology-related knowledge, skills, and attitudes (KSA) of teachers (section 2.1 and 2.2), present how these KSA are currently assessed (section 2.3), and outline ways of fostering them among pre- and in-service teachers (section 2.4). Section 3 explains how the eight articles in this special issue contribute to the knowledge base in these three areas. Section 4 provides an outlook on how technology-related KSA might change in the future due to new technological developments, especially artificial intelligence.

2. Technology-related KSA

2.1. Knowledge and skills

Successful problem solvers possess well-organized and flexible reservoirs of knowledge that they can apply within various contexts (Ericsson & Lehmann, 1996). Drawing on the work of Shulman (1986; 1987), Park and Oliver (2008) identified in their literature review four communalities of teachers' professional knowledge: pedagogical knowledge (PK), content knowledge (CK), pedagogical content knowledge (PCK), and knowledge of context. Mishra and Koehler (2006) added technological knowledge (TK) as a further type (Koehler & Mishra, 2009; Mishra & Koehler, 2006). Their so-called TPACK framework (see Fig. 1) has gained broad attention among researchers (Harris, Phillips, Koehler, & Rosenberg, 2017; Hew, Lan, Tang, Jia, & Lo, 2019; Petko, 2020; Saubern, Henderson, Heinrich, & Redmond, 2020; Voogt, Fisser, Pareja Roblin, Tondeur, & van Braak, 2013); most of the contributions to this special issue use it as a theoretical background. The core that emerges from interaction between CK, PK, and TK (Koehler et al., 2013) is TPACK (see Fig. 1).

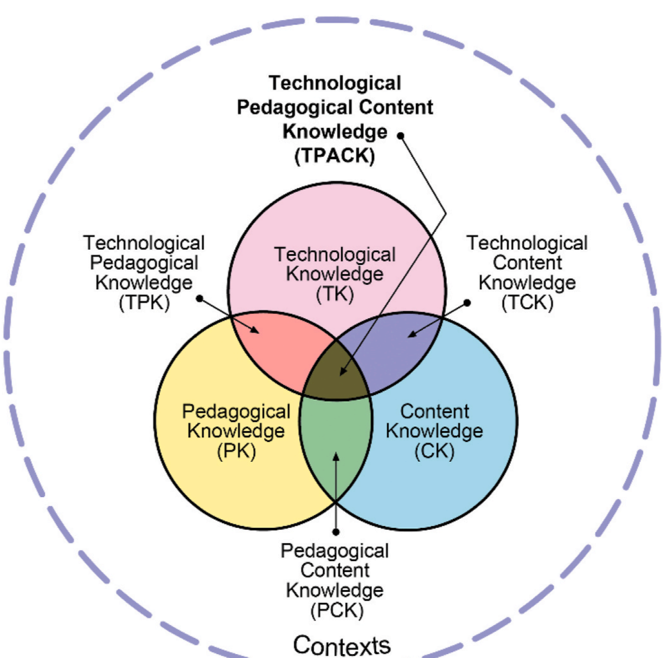


Fig. 1. TPACK framework. Reproduced by permission of the publisher © 2012 by tpack.org.

As the name implies, TPACK may refer to knowledge; however, researchers also conceptualized the components of the TPACK framework as skills (i.e., the application of knowledge) or competence (Willermark, 2018). For a discussion on this, see Sailer et al. (2020, this issue) and Wekerle and Kollar (2020, this issue).

Although the TPACK framework is well-established, it has been challenged (Angeli & Valanides, 2009; Graham, 2011): The building blocks PK, CK, and TK are insufficiently (theoretically) conceptualized. In particular, it often remains unclear what kind of technology TK comprises. The range could be broad, e.g., from chalk boards to social robots. Besides the underlying constructs themselves, their relationship with each other is an issue (Petko, 2020). This relationship could be integrative or transformative. The integrative view implies that all constructs directly contribute to TPACK, e.g., an increase in TK directly yields an increase in TPACK. The transformative view posits that the influence of CK, PK, and TK on TPACK is fully mediated by TPK, TCK, and PCK (see Fig. 1). Based on a sample of pre-service teachers, Schmid, Brianza, and Petko (2020) recently provided evidence in favor of the transformative view. However, using a sample of in-service teachers, the findings of Koh, Chai, and Tsai (2013) may, in part, point to an integrative view because TK and PK are directly associated with TPACK. A further criticism of the TPACK framework is its ability to predict meaningful outcomes (Graham, 2011). Available studies in general refer to (self-reported) use (Farjon, Smits, & Voogt, 2019; Guggemos & Seufert, 2020, this issue; Schmid, Brianza, & Petko, 2020, this issue). However, in the end, the desired outcome would be gains in student learning. In other contexts, studies that rely on student learning are available. For instance, Baumert et al. (2010) showed in a longitudinal study that the objectively measured PCK of teachers can explain class variance in mathematics achievement reasonably well ($R^2 = 39\%$).

2.2. Attitudes

Besides professional knowledge and skills, teachers' attitudes, especially the beliefs that form such attitudes (Instefjord & Munthe, 2017), have received much attention. Empirical evidence lends support to the important role of beliefs in the process of technology integration (Cheng & Xie, 2018; Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012; Petko, 2012; Tondeur, van Braak, Ertmer, & Ottenbreit-Leftwich, 2017). Unsurprisingly, attitudes also play an important role in predicting the adoption of technology by teachers. Following the theory of planned behavior (Ajzen, 1991), the attitude towards a behavior is one of three predictors for behavioral intention. Moreover, the will, skill, tool model implies attitudes are a predictor for the actual use of technology (Knezek & Christensen, 2016). Indeed, Scherer and Teo (2019) identified in their meta-analysis attitudes as a significant predictor for teachers' intention to use technology; however, they did not consider the actual use, which might be problematic (Nistor, 2014). Scherer, Tondeur, Siddiq, and Baran (2018) showed that attitudes towards technology can be an important predictor for the level of TPACK in pre-service teachers.

2.3. Assessment of technology-related KSA

Quantitative research about technology integration depends on instruments to measure the constructs of interest; the validity of the findings heavily depends on the quality of the used instruments. Moreover, for professional development, instruments are necessary to identify potentials for improvement (formative assessment). In general, assessment instruments in the realm of teachers' technology-related KSA can be separated into self-assessment and external (objective) instruments (Kaplon-Schilis & Lyublinskaya, 2020). Although the use of self-assessment instruments has been challenged (Lachner, Backfisch, & Stürmer, 2019), they are regularly used to capture technology-related KSA. Various self-assessment instruments are available (e.g., Schmid et al., 2020; Valtonen et al., 2017; for an overview see Voogt et al., 2013). Generally speaking, self-assessments may not be inferior to objective measures (Conway & Lance, 2010). Rather, the validity of an assessment can only be evaluated with respect to its purpose (AERA, APA, & NCME, 2014). Scherer, Tondeur, and Siddiq (2017) pointed out five advantages of self-assessment instruments. First, they are a cost-efficient, reliable, and valid indicator for teachers' self-efficacy beliefs. Second, they are important predictors for teachers' intention to use technology. Third, they are geared towards future behavior. Fourth, self-efficacy beliefs correlate with the quality of instruction and favorable educational outcomes. Fifth, self-efficacy beliefs are a desirable outcome of the education and professional development of teachers. In summary, self-reports and objective measures might capture different constructs (self-efficacy beliefs vs. performance) that are both important; hence, they may be regarded as complementary (Drummond & Sweeney, 2017).

For a valid assessment, both curriculum and instruction have to be taken into account (Pellegrino, DiBello, & Goldman, 2016). The curriculum specifies the kind of knowledge (pre-service) teachers are expected to possess. Hence, if teachers' technology-related KSA are to be assessed, the pertinent standards have to be considered. For instance, Tondeur, Aesaert, et al. (2017) reviewed various standards to develop an instrument to measure the technological skills of pre-service teachers. Sailer et al. (2020, this issue) and Rubach et al. (2020, this issue) also considered the pertinent curricula for developing self-assessment instruments. Such an approach, where the facets that constitute technology-related knowledge or skills are clearly pointed out, may also be conducive for identifying potential aspects of professional development. Besides the necessity for clearly specifying the constructs' structure, proficiency levels stating at what level teachers are expected to possess these technology-related KSA are important. To form such proficiency levels, Saubern, Urbach, Koehler, and Phillips (2020) relied on item response theory and developed a model of teachers' TPACK confidence that comprises five levels. By means of such proficiency levels, the results of an assessment can be interpreted by referring to clearly specified criteria.

Currently, there seems to be much room for improvement in the practice of assessment (Petko, 2020). To move forwards, it may be helpful to consider research about assessing the professional development of teachers in general, e.g., Blömeke and Delaney (2012). For PK, Voss, Kunter, and Baumert (2011) developed a model comprising five facets, e.g., classroom management. As Nistor (2014) argued in the realm of technology acceptance, it may be necessary to consider in a better manner the complexity and dimensionality of the underlying constructs. Such a claim might be in line with Saubern et al. (2020, p. 6) who stated it may be important "to understand the knowledge that teachers need to use technology effectively for teaching and learning", i.e., what are the building blocks that constitute technology-related knowledge. Following the call for more elaborate instruments, Rubach et al. (2020, this issue), for instance, use a six-dimensional model to develop a self-assessment instrument that measures TK. To form proficiency levels, the approach of Hartig, Frey, Nold, and Klieme (2012) may be helpful.

2.4. *Fostering technology-related knowledge, skills, and attitudes*

Adopting technology for learning purposes is a complex endeavor (Straub, 2009). We cannot expect prospective teachers to possess technology-related knowledge and skills simply because they grew up with digital technology (for more on the myth of digital natives, see Kirschner & Bruyckere, 2017). Rather, training and professional development might be necessary. Based on a systematic review of qualitative studies, Tondeur et al. (2012) developed a Synthesis of Qualitative Evidence (SQD) model for preparing pre-service teachers to include technology in their classroom practices. They identified six themes on the micro level concerning the preparation of pre-service teachers. A questionnaire that operationalizes these six themes is available (Tondeur, van Braak, Siddiq, & Scherer, 2016). Moreover, there is evidence for a positive association between the perceived occurrence of the SQD strategies and TPACK (Tondeur, Scherer, Siddiq, & Baran, 2020). Table 1 summarizes the themes of the SQD model.

The SQD model may not only be helpful for training pre-service teachers, but also for professional development in general. Yurtseven Avci, O'Dwyer, and Lawson (2020) reviewed thirty-two studies about professional development in the realm of technology in education. Overall, they identified categories that are consistent with the SQD model. For instance, they suggest that skilled teachers with a highly positive attitude towards technology could act as coaches or mentors to support their colleagues. What might be specific to in-service teachers could be learning communities or communities of practice that play an important role in the professional development of teachers (Lawless & Pellegrino, 2007; Tseng & Kuo, 2014).

3. Contributions in this special issue

3.1. *Overview*

The eight articles and the afterword in this special issue aim at fostering a better understanding of the technology-related KSA of pre- and in-service teachers; Table 2 provides an overview.

3.2. *Specifying technology-related KSA*

Guggemos and Seufert (2020, this issue) use the TPACK framework as the theoretical background to predict teachers' use of technology as a means and as the content of instruction (teaching with and about technology). They rely on structural equation modeling to test their hypotheses. For the critical conceptualization of 'technology', they utilized the DigComp 2.1 framework of the European Union (Carretero, Vuorikari, & Punie, 2017). Moreover, they consider a new type of knowledge—technological collaboration knowledge (TCoK). Interaction and collaboration is important in knowledge acquisition in general (Chi & Wylie, 2014) and especially so in the professional development of teachers (Vangrieken, Meredith, Packer, & Kyndt, 2017). However, collaboration is seldom taken into account in TPACK research. Most available studies that rely on the TPACK framework do not consider the collaborative aspect. By integrating TCoK, Guggemos and Seufert (2020, this issue) may be in line with Petko (2020), who argues that for theory development, the TPACK framework could be extended without changing its core.

Hämäläinen et al. (2020, this issue) utilize data from large-scale studies (PIAAC and TALIS) and conduct secondary analyses. The

Table 1
Micro-level themes of the SQD model.

Theme	Manifestation in teacher training
Role model	Pedagogical meaningful use of technology is embedded in all kinds of activities, e.g., in lectures and seminars
Reflection	Reflection and discussion on the use of technology is an integral part
Instructional design	Students receive help in preparing lessons that include technology
Collaboration	Students have opportunities to work together with fellow students, supporting each other, and sharing experiences
Authentic experiences	Students receive opportunities to test themselves using technology in the classroom (in internships)
Feedback	Students receive feedback about their use of technology and about further improvements

Note. The SQD model as presented by Tondeur et al. (2012; 2016).

Table 2
Overview of the contributions.

No.	Authors	Context	Sample size
Specifying technology-related KSA			
1	Guggemos and Seufert	In-service, Switzerland, VET schools	212
2	Hämäläinen, Nissinen, Taajamo, and Lämsä	In-service, 14 countries, various school levels	67,102
3	Valtonen, Hoang, Sointu, Näykki, Mäkitalo, Kukkonen, Virtanen, Järvelä, and Häkkinen	Pre-service, Finland, high schools, three-wave longitudinal	704
Assessing technology-related KSA			
4	Schmid, Brianza, and Petko	Pre-service, Switzerland, middle schools	173
5	Sailer, Stadler, Schultz-Pernice, Franke, Schöffmann, Paniotova, Husagic, and Fischer	Pre- and in-service, Germany, primary and high schools	90
6	Rubach, Lazarides, and Quast	In-service, Germany, primary and high schools	449
Fostering technology-related KSA			
7	Wekerle and Kollar	Pre-service, Germany, high schools	252
8	Tondeur, Howard, and Yang	Pre-service, Belgium, high schools	931
Afterword			
9	Mishra and Warr	conceptual	n/a

authors take advantage of the very large sample sizes of these studies and investigate to what extent teachers possess technology-related KSA. Personal and contextual factors are considered as well. While there is notable variance in teachers' knowledge and skills, their attitudes show less variance as they generally recognize the importance of teaching with digital technologies. However, despite the fact that teachers, in general, seem to recognize the importance of technology use, some teachers do not rate digital skills as very important for their work. On one hand, there is a group of teaching professionals which believes their digital skills are adequate and this group does indeed have advanced digital skills. On the other, another group of teaching professionals believes their skills are adequate for their work but teachers from this group have a poor grasp of digital skills. Based on their findings, the authors present implications for professional development.

In general, technology-related KSA can be embedded within the wider context of 21st-century skills (Kirschner & Stoyanov, 2020; van Laar, van Deursen, van Dijk, & Haan, 2017; Voogt & Roblin, 2012). In a similar vein, *Valtonen* et al. (2020, this issue) investigated the development of 21st-century skills of pre-service teachers at three Finnish universities using latent growth curve modeling. The TPACK framework and the theory of planned behavior acts as the theoretical background for operationalizing technology-related knowledge and self-efficacy. Such longitudinal designs may be important for obtaining a better understanding of developmental processes (Petko, 2020). Interestingly, TPK and technology self-efficacy showed significant growth, whereas other 21st-century skills, such as negotiation or team leadership, remained rather stable over the three-year time frame.

3.3. Assessing technology-related KSA

Schmid, Brianza, and Petko (2020, this issue) address an important question in TPACK research, namely, the validity of self-assessments. To this end, they investigate how those constructs (see Fig. 1) measured with self-assessment instruments relate to planned technology use in lessons. To capture the planned technology use, pre-service teachers were asked to design a lesson for a topic of their choice; no explicit prompt for including the use of technology was given. Moreover, the authors carry out a cluster analysis to identify TPACK profiles among the pre-service teachers in their sample that could predict planned technology use in lessons. In this vein, they distinguish between use of technology by the teacher and use of technology by the students. However, neither the constructs of the TPACK framework nor the identified profiles in the cluster analysis are associated with technology use in lessons. The authors discuss these remarkable findings.

Sailer et al. (2020, this issue) present and validate by means of confirmatory factor analysis a self-assessment instrument for teachers' technology-related skills and attitudes. Their work goes beyond available assessment instruments in many ways. First, they use scenarios to make the type and circumstances of technology use clear to test takers. Second, they address skills, whereas many studies focus on knowledge. Third, they take pertinent curricula into account and consider instructional practices, which may contribute to the validity of the instrument (Pellegrino et al., 2016). Fourth, they consider the use of technology across all kinds of teaching activities, e.g., planning and evaluating, instead of solely teaching in the classroom. Fifth, when validating their instrument, they not only take the frequency but also the kind of technology use into account (Chi & Wylie, 2014). The developed instrument is freely available.

Rubach, Lazarides, and Quast (2020, this issue) also present and validate by means of confirmatory factor analysis a self-assessment instrument. The construct of interest is ICT competence beliefs, which may be regarded as TK based on the TPACK framework. In line with the call for more elaborate instruments, they rely on a six-dimensional model that is based on pertinent standards, e.g., DigComp (Ferrari, 2013). The facets are 'Information and data literacy', 'Communication and collaboration', 'Digital content creation', 'Safety and security', 'Problem solving', and 'Analyzing and reflecting'. The authors provide evidence for the structure of their model, as well as for convergent and discriminant validity. The validated instrument is part of the article.

3.4. Fostering technology-related KSA

In line with the call of Petko (2020) for more rigor in research designs, *Wekerle and Kollar* (2020, this issue) utilize an experimental design and objective measures. The TPACK framework acts as the theoretical background. However, instead of knowledge they focus on situation-specific skills (Blömeke, Gustafsson, & Shavelson, 2015) in the form of professional vision. In their 2×2 between-subjects design they vary the factor mapping (reading vs. mapping) and worked examples (problem solving vs. worked examples). The students read three texts covering PK-related content. In the mapping conditions students were further instructed to create a combined mind map. They then worked with written cases of classroom lessons that included technology-enhanced teacher and student actions. They either received worked examples on how to reason regarding these cases and were prompted to explain them (worked examples), or they were asked to reason with these cases without further instruction (problem solving). Wekerle and Kollar conclude by suggesting that learning by mapping, if carried out properly, might have a positive influence on declarative knowledge. Students who studied worked examples (instead of solving problems) showed better noticing, explanation, and solution skills.

As described in section 2.4, the SQD is a model for training pre-service teachers. *Tondeur* et al. (2020, this issue) aim at increasing the usefulness of this model. To this end, they utilize association rules analysis as a data mining technique (García, Romero, Ventura, & Castro, 2011). Here, rules are formulated as if-then statements, e.g., if A and B, then C and D. Each rule includes at least one antecedent (A) and one consequent (C). In the case of Tondeur et al. the rules are formed based on answers to an SQD questionnaire. Furthermore, they distinguished between teachers with negative and positive attitudes towards technology. They found remarkably different association patterns. Pre-service teachers with negative attitudes showed an emphasis on feedback, and students with positive attitudes towards collaboration. This different structure of associations could be regarded as further evidence for the important role of attitudes towards technology for understanding technology integration. Moreover, the authors discuss the implications for more personalized training of pre-service teachers.

3.5. Afterword

The contributions in this special issue are found on the micro-level, i. e., the focus is on individual teachers. In the afterword, *Mishra and Parr* (2020, this issue) widen the perspective. For an in-depth understanding of technology integration, they argue that specific systems and cultures of practice have to be considered. With this in mind, they present the five spaces framework. This model might also reveal promising avenues for further research which may contribute to a better understanding of technology integration.

4. Outlook

In terms of digitalization, it may be helpful to differentiate between two waves (Wahlster, 2017). The first wave has been driven by the digitization of data and the internet. The studies in this special issue are (due to their quantitative empirical nature) part of the first wave. The second wave is characterized by technologies like artificial intelligence (AI), which focus on understanding data in order to utilize them for various purposes. In education, the potential of AI has not been fully exploited (Luckin, Holmes, Griffiths, & Forcier, 2016). Recent publications and major international funding schemes point to the perceived importance of AI in education (Hasse, Cortesi, Lombana, & Gasser, 2019; Lytras, Sarirete, & Damiani, 2020; Stanford University, 2016; Yueh & Chiang, 2020). Concerning the role of AI in education, Holmes, Bialik, and Fadel (2019) pointed to a crucial point: "Whether we welcome it or not, AI is increasingly being used widely across education and learning contexts. We can either leave it to others—the computer scientists, AI engineers and big tech companies—to decide how artificial intelligence in education unfolds, or we can … adopt a critical stance, to help ensure that the introduction of AI into education reaches its potential and has positive outcomes for all." (Holmes et al., 2019, p. 179).

For a better understanding of the impact of AI on professionals in education, e.g., teachers, the concept of a*ugmentation strategies* by Davenport and Kirby (2016) could be helpful. This generic framework describes how the relationship between humans and smart machines (driven by AI) might unfold in the future. In the following section, we will point out what augmentation strategies could evolve in education. On the one hand, new specialized skill profiles may be needed (e.g., educational data scientists, the 'step in' strategy) and, on the other hand, the skill profiles of existing roles in teaching and learning might change (e.g., smart machines support teachers during instruction, the 'step aside' strategy). Moreover, a 'step forward' strategy is possible (e.g., contributing to the development of AI-based learning solutions). A 'step up' strategy is especially important for education and school managers. It deals with important questions of data protection and the use of student data in an ethical way because, as with every technology, AI also poses dangers. "[We] are excited by what AI has to offer teaching and learning … but we are also very cautious. We have seen an extraordinary range of AIED approaches … and some amazing future AIED possibilities … However, we have also identified a range of critical issues that need to be addressed before AI becomes an acceptable integral part of everyday learning." (Holmes et al., 2019, p. 179).

The augmentation strategies show how the role of professionals in education could look like within the context of a second wave of digitalization, as well as which KSA might be necessary depending on the pursued strategy. However, the question that remains is how can such a transformation process be integrated into the daily business of schools? In organization theory, ambidexterity, as a so-called two-handed leadership approach, has long been a much noticed concept (O'Reilly & Tushman, 2008) . It addresses the ability to exploit existing opportunities (left hand) while at the same time exploring new ones (right hand). Accompanied by this is the question of how exploitation (optimizing the existing) and exploration (exploring the new) can be integrated into an organization such as a school. Principals and teachers are currently faced with the challenge of integrating exploration alongside their everyday business (exploitation). There is a risk that investments made today in certain digital skills will be outdated in five years' time. New interactions between human beings and smart machines raise new fundamental questions that cannot yet be answered comprehensively by the education and professional development of teachers. Therefore, it might be useful to better understand and further investigate the concept of augmented work and augmentation strategies for the teaching profession in order to identify the technology-related KSA of teachers (Seufert, Guggemos, & Sonderegger, 2020).

The second wave of digitalization does not mean that less teachers are needed (Dillenbourg, 2016). Human expectations towards smart machines are certainly less sensitive than those towards human teachers. The augmented work of the teaching profession has to be researched, as well as how teachers might engage with smart machines within their teaching environment. Dillenbourg (2013) argues for the orchestration of learning—the real-time management of classroom activities—across the classroom ecosystem. In the future, this orchestration process could include smart machines.

The direction of thought has been deliberately redirected: not always planning the next few years based on the current situation but, conversely, designing scenarios for the future and, from this perspective, exploring possible areas of development on how learning and teaching in schools might look in the future. Teachers should be enabled and supported to sit in the driver seat in order to shape their schools during this current major transitional phase—to quote Kay's (1971) well-known remark: "The best way to predict the future is to invent it."

CRediT authorship contribution statement

Sabine Seufert: Conceptualization, Writing - original draft, Writing - review & editing, Supervision. **Josef Guggemos:** Conceptualization, Writing - original draft, Writing - review & editing, Project administration. **Michael Sailer:** Conceptualization, Writing - original draft, Writing - review & editing, Project administration.

Acknowledgements

The Guest Editors are thankful for the ongoing support received from Prof. Paul A. Kirschner, the Commissioning Guest Editor, and the reviewers who ensured the quality of this special issue. Moreover, we are indebted to Prof. Björn de Koning who acted as the Editor for two manuscripts.

References

AERA, A. P. A., & NCME. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education, 52*(1), 154–168. https://doi.org/10.1016/j.compedu.2008.07.006

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM, 44*(3), 223–247. https://doi.org/10.1007/s11858-012-0429-7

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie, 223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Carretero, S., Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. Luxembourg: European Union (EU). https://doi.org/10.2760/38842

Cheng, S.-L., & Xie, K. (2018). The relations among teacher value beliefs, personal characteristics, and TPACK in intervention and non-intervention settings. *Teaching and Teacher Education, 74*, 98–113. https://doi.org/10.1016/j.tate.2018.04.014

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology, 25*(3), 325–334. https://doi.org/10.1007/s10869-010-9181-6

Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines*. Harper Business.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education, 69*, 485–492. https://doi.org/10.1016/j.compedu.2013.04.013

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education, 26*(2), 544–560. https://doi.org/10.1007/s40593-016-0106-z

Drummond, A., & Sweeney, T. (2017). Can an objective measure of technological pedagogical content knowledge (TPACK) supplement existing TPACK measures? *British Journal of Educational Technology, 48*(4), 928–939. https://doi.org/10.1111/bjet.12473

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education, 59*(2), 423–435. https://doi.org/10.1016/j.compedu.2012.02.001

Farjon, D., Smits, A., & Voogt, J. M. (2019). Technology integration of pre-service teachers explained by attitudes and beliefs, competency, access, and experience. *Computers & Education, 130*, 81–93. https://doi.org/10.1016/j.compedu.2018.11.010

Ferrari, A. (2013). *Digcomp: A framework for developing and understanding digital competence in europe*. Publications Office.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA). https://www.iea.nl/publications/study-reports/preparing-life-digital-world.

García, E., Romero, C., Ventura, S., & Castro, C. de (2011). A collaborative educational association rule mining tool. *The Internet and Higher Education, 14*(2), 77–88. https://doi.org/10.1016/j.iheduc.2010.07.006

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education, 57*(3), 1953–1960. https://doi.org/10.1016/j.compedu.2011.04.010

Harris, J. B., Phillips, M., Koehler, M. J., & Rosenberg, J. M. (2017). TPCK/TPACK research and development: Past, present, and future directions. *Australasian Journal of Educational Technology, 33*(3). https://doi.org/10.14742/ajet.3907

Hartig, J., Frey, A., Nold, G., & Klieme, E. (2012). An application of explanatory item response modeling for model-based proficiency scaling. *Educational and Psychological Measurement, 72*(4), 665–686. https://doi.org/10.1177/0013164411430707

Hasse, A., Cortesi, S. C., Lombana, A., & Gasser, U. (2019). Youth and artificial intelligence: Where we stand. In *SSRN electronic journal*. Advance online publication. https://doi.org/10.2139/ssrn.3385718.

Hew, K. F., Lan, M., Tang, Y., Jia, C., & Lo, C. K. (2019). Where is the "theory" within the field of educational technology research? *British Journal of Educational Technology, 50*(3), 956–971. https://doi.org/10.1111/bjet.12770

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). *The difference between emergency remote teaching and online learning*. https://medicine.hofstra.edu/pdf/faculty/facdev/facdev-article.pdf.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. The Center for Curriculum Redesign.

Instefjord, E. J., & Munthe, E. (2017). Educating digitally competent teachers: A study of integration of professional digital competence in teacher education. *Teaching and Teacher Education, 67*, 37–45. https://doi.org/10.1016/j.tate.2017.05.016

J-PAL. (2019). *Will technology transform education for the better?*. https://www.povertyactionlab.org/publication/will-technology-transform-education-better.

Kaplon-Schilis, A., & Lyublinskaya, I. (2020). Analysis of relationship between five domains of TPACK framework: TK, PK, CK math, CK science, and TPACK of pre-service special education teachers. *Technology, Knowledge and Learning, 25*(1), 25–43. https://doi.org/10.1007/s10758-019-09404-x

Kay, A. (1971). *The best way to predict the future is to invent it*. https://en.wikiquote.org/wiki/Alan_Kay.

Kirschner, P. A. (2015). Do we need teachers as designers of technology enhanced learning? *Instructional Science, 43*(2), 309–322. https://doi.org/10.1007/s11251-015-9346-9

Kirschner, P. A., & Bruyckere, P. de (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education, 67*, 135–142. https://doi.org/10.1016/j.tate.2017.06.001

Kirschner, P. A., & Stoyanov, S. (2020). Educating youth for nonexistent/not yet existing professions. *Educational Policy, 34*(3), 477–517. https://doi.org/10.1177/0895904818802086

Kirschner, P. A., & van Merrienboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist, 48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Knezek, G., & Christensen, R. (2016). Extending the will, skill, tool model of technology integration: Adding pedagogy as a new model construct. *Journal of Computing in Higher Education, 28*(3), 307–325. https://doi.org/10.1007/s12528-016-9120-2

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.

Koehler, M. J., Mishra, P., & Cain, W. (2013). What is technological pedagogical content knowledge (TPACK)? *Journal of Education, 193*(3), 13–19. https://doi.org/10.1177/002205741319300303

Koh, J. H. L., Chai, C. S., & Tsai, C.-C. (2013). Examining practicing teachers' perceptions of technological pedagogical content knowledge (TPACK) pathways: A structural equation modeling approach. *Instructional Science, 41*(4), 793–809. https://doi.org/10.1007/s11251-012-9249-y

Lachner, A., Backfisch, I., & Stürmer, K. (2019). A test-based approach of modeling and measuring Technological Pedagogical Knowledge. *Computers & Education, 142*, 103645. https://doi.org/10.1016/j.compedu.2019.103645

Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research, 77*(4), 575–614. https://doi.org/10.3102/0034654307309921

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Lytras, M., Sarirete, A., & Damiani, E. (2020). Technology-enhanced learning research in higher education: A transformative education primer. *Computers in Human Behavior, 109*, 106350. https://doi.org/10.1016/j.chb.2020.106350

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research & Development, 50*(3), 43–59. https://doi.org/10.1007/BF02505024

Mishra, P., & Koehler, M. J. (2006). Technological edagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Nistor, N. (2014). When technology acceptance models won't work: Non-significant intention-behavior effects. *Computers in Human Behavior, 34*, 299–300. https://doi.org/10.1016/j.chb.2014.02.052

OECD. (2015). *Students, computers and learning*. Advance online publication. https://doi.org/10.1787/9789264239555-en

O'Reilly, C. A., & Tushman, M. L. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior, 28*, 185–206. https://doi.org/10.1016/J.RIOB.2008.06.002

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59–81. https://doi.org/10.1080/00461520.2016.1145550

Petko, D. (2012). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the 'will, skill, tool' model and integrating teachers' constructivist orientations. *Computers & Education, 58*(4), 1351–1359. https://doi.org/10.1016/j.compedu.2011.12.013

Petko, D. (2020). Quo vadis TPACK? Scouting the road ahead. *EdMedia + Innovate Learning, 23*, 1277–1286.

Saubern, R., Henderson, M., Heinrich, E., & Redmond, P. (2020). Tpack – time to reboot? *Australasian Journal of Educational Technology, 36*(3), 1–9. https://doi.org/10.14742/ajet.6378

Saubern, R., Urbach, D., Koehler, M., & Phillips, M. (2020). Describing increasing proficiency in teachers' knowledge of the effective use of digital technology. *Computers & Education, 147*, 103784. https://doi.org/10.1016/j.compedu.2019.103784

Scherer, R., & Teo, T. (2019). Unpacking teachers' intentions to integrate technology: A meta-analysis. *Educational Research Review, 27*, 90–109. https://doi.org/10.1016/j.edurev.2019.03.001

Scherer, R., Tondeur, J., & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education, 112*, 1–17. https://doi.org/10.1016/j.compedu.2017.04.012

Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior, 80*, 67–80. https://doi.org/10.1016/j.chb.2017.11.003

Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education, 157*, 103967. https://doi.org/10.1016/j.compedu.2020.103967

Seufert, S., Guggemos, J., & Sonderegger, S. (2020). Digitale Transformation der Hochschullehre: Augmentationsstrategien für den Einsatz von Data Analytics und Künstlicher Intelligenz [Digital transformation in higher education: Augmentation strategies for the use of data analytics and artificial intelligence]. *ZFHE, 15*(1), 81–101. https://doi.org/10.3217/zfhe-15-01/05

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Stanford University. (2016). *Artificial intelligence and life in 2030. One hundred year study on artificial intelligence: Report of the 2015-2016 study panel*. Author https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf.

Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. *Review of Educational Research, 79*(2), 625–649. https://doi.org/10.3102/0034654308325896

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research & Development, 68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning. *Review of Educational Research, 81*(1), 4–28. https://doi.org/10.3102/0034654310393361

Tondeur, J., Aesaert, K., Pynoo, B., van Braak, J., Fraeyman, N., & Erstad, O. (2017). Developing a validated instrument to measure preservice teachers' ICT competencies: Meeting the demands of the 21st century. *British Journal of Educational Technology, 48*(2), 462–472. https://doi.org/10.1111/bjet.12380

Tondeur, J., Scherer, R., Siddiq, F., & Baran, E. (2020). Enhancing pre-service teachers' technological pedagogical content knowledge (TPACK): A mixed-method study. *Educational Technology Research & Development, 68*(1), 319–343. https://doi.org/10.1007/s11423-019-09692-1

Tondeur, J., van Braak, J., Ertmer, P. A., & Ottenbreit-Leftwich, A. (2017). Understanding the relationship between teachers' pedagogical beliefs and technology use in education: A systematic review of qualitative evidence. *Educational Technology Research & Development, 65*(3), 555–575. https://doi.org/10.1007/s11423-016-9481-2

Tondeur, J., van Braak, J., Sang, G., Voogt, J. M., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education, 59*(1), 134–144. https://doi.org/10.1016/j.compedu.2011.10.009

Tondeur, J., van Braak, J., Siddiq, F., & Scherer, R. (2016). Time for a new approach to prepare future teachers for educational technology use: Its meaning and measurement. *Computers & Education, 94*, 134–150. https://doi.org/10.1016/j.compedu.2015.11.009

Tseng, F.-C., & Kuo, F.-Y. (2014). A study of social participation and knowledge sharing in the teachers' online professional community of practice. *Computers & Education, 72*, 37–47. https://doi.org/10.1016/j.compedu.2013.10.005

Valtonen, T., Sointu, E., Kukkonen, J., Kontkanen, S., Lambert, M. C., & Mäkitalo-Siegl, K. (2017). TPACK updated to measure pre-service teachers' twenty-first century skills. *Australasian Journal of Educational Technology, 33*(3). https://doi.org/10.14742/ajet.3518

Van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & Haan, J. de (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577–588. https://doi.org/10.1016/j.chb.2017.03.010

Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education, 61*, 47–59. https://doi.org/10.1016/j.tate.2016.10.001

Voogt, J. M., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge - a review of the literature. *Journal of Computer Assisted Learning, 29*(2), 109–121. https://doi.org/10.1111/j.1365-2729.2012.00487.x

Voogt, J. M., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21 st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies, 44*(3), 299–321. https://doi.org/10.1080/00220272.2012.668938

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology, 103*(4), 952–969. https://doi.org/10.1037/a0025125

Wahlster, W. (2017). Artificial intelligence as the driver for the second wave of digitalization. *IM+io*, (2), 10–13.

Willermark, S. (2018). Technological pedagogical and content knowledge: A review of empirical studies published from 2011 to 2016. *Journal of Educational Computing Research, 56*(3), 315–343. https://doi.org/10.1177/0735633117713114

Yueh, H.-P., & Chiang, F.-K. (2020). Editorial: AI and robotics in reshaping the dynamics of learning. In *British journal of educational technology*. Advance online publication. https://doi.org/10.1111/bjet.13017.

Yurtseven Avci, Z., O'Dwyer, L. M., & Lawson, J. (2020). Designing effective professional development for technology integration in schools. *Journal of Computer Assisted Learning, 36*(2), 160–177. https://doi.org/10.1111/jcal.12394

Studie 10:

Guggemos, J., & Seufert, S. (2021). Teaching with and teaching about technology – Evidence for professional development of in-service teachers. *Computers in Human Behavior*, 115, 106613.

Computers in Human Behavior 115 (2021) 106613

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Teaching with and teaching about technology – Evidence for professional development of in-service teachers

Josef Guggemos *, Sabine Seufert

School of Management, Institute of Business Education and Educational Management, University of St.Gallen, St. Jakob-Strasse 21, 9000, St. Gallen, Switzerland

A R T I C L E I N F O

Keywords:
Digital transformation
Teacher professional development
TPACK
Will, Skill, Tool model
Teacher collaboration
DigComp framework

A B S T R A C T

The digital transformation has implications for how and what to teach. For the purpose of professional development, the paper at hand presents a conceptual framework for predicting the use of technology as a means and as a content of instruction. It is informed by the TPACK framework and the 'will, skill, tool' model. The predictors are Technological Knowledge (TK), Technological Pedagogical Knowledge (TPK), Technological Pedagogical Content Knowledge (TPACK), Technological Collaboration Knowledge (TCoK), and Attitudes. These constructs are measured by newly developed self-assessment instruments. Structural equation modeling using a sample of 212 in-service teachers from commercial schools in German-speaking Switzerland lend support to the soundness of the measurement instrument and the conceptual framework. Overall, 36% of the variance of the use of technology as a means and 45% of the variance for the use as the content of instruction can be explained. Mediation and multigroup analyses, a finite-mixture segmentation, comparisons of competing models, and factor score regression yielded evidence for the robustness of the conceptual framework.

1. Introduction

Consensus may be that teachers play a crucial role in the process of integrating technology into instruction (Dillenbourg, 2013; Kirschner, 2015; Lawless & Pellegrino, 2007; OECD, 2015). Besides this, they might also be required to foster technology-related skills among students (Claro et al., 2018; Redecker, 2017; Siddiq, Scherer, & Tondeur, 2016). The 2018 Information and Computer Literacy Study (ICILS) summarizes this dual role as "Teaching with and about information and communications technologies" (Fraillon, Ainley, Schulz, Friedman, & Duckworth, 2019, p. 175). In the German-speaking context, this two-sidedness is represented in the construct (digital) media competencies that comprises media didactics and media education (Tiede, Grafe, & Hobbs, 2015). On the one hand, teachers are supposed to use technology in their instruction in a way that is conducive to achieving meaningful pedagogical goals; on the other, teachers may be supposed to integrate new content into their instruction or change the instructional focus due to the digital transformation. Information and communication technology (ICT) literacy, for instance, may become increasingly important for participating in society and for being successful in the workplace (Fraillon et al., 2019). Instruction in this regard may be necessary because students are not by nature proficient ICT users. The idea of digital natives who use technology competently without specific training seems to be a myth (Kirschner & Bruyckere, 2017).

Teachers' professional knowledge may become the focus point because it has been shown to be an important predictor for instructional quality (Baumert et al., 2010). Generic frameworks that address necessary knowledge in the context of digital transformation are available (Oberländer, Beinicke, & Bipp, 2020; Van Laar, van Deursen, van Dijk, & Haan, 2017). In Europe, the DigComp framework, issued by the European Commission, may be influential (Carretero, Vuorikari, & Punie, 2017). It comprises five facets: information and data literacy, digital collaboration and communication, digital content creation, ensuring digital safety, and solving technical problems. Based on the generic DigComp, the DigCompEdu framework adds further facets necessary for educators (Redecker, 2017). The DigCompOrg framework complements DigComp by addressing the integration of digital learning on an organizational level (Kampylis, Punie, & Devine, 2015). In the research-oriented literature, the TPACK framework (Koehler & Mishra, 2009; Mishra & Koehler, 2006), in particular, has gained broad attention (Petko, 2020; Voogt, Fisser, Pareja Roblin, Tondeur, & van Braak, 2013). Drawing from Shulman (1986, 1987), Mishra and Koehler (2006) added a further facet to content knowledge (CK) and pedagogical knowledge (PK): technological knowledge (TK). For the successful integration of

* Corresponding author.
E-mail address: josef.guggemos@unisg.ch (J. Guggemos).

https://doi.org/10.1016/j.chb.2020.106613
Received 28 May 2020; Received in revised form 14 October 2020; Accepted 25 October 2020
Available online 3 November 2020

technology into a specific subject, TPACK, a combination of CK, PK, and TK, is necessary (Koehler & Mishra, 2009). Although widespread, the TPACK framework has been challenged (Angeli & Valanides, 2009). Graham (2011) argued that the delineation between the three elements is sometimes fuzzy, their relationship with each other needs clarification, and its use for predicting meaningful outcomes is doubtful. Moreover, TPACK addresses the integration of technology into instruction. Whether it could also contribute to the understanding of integrating technology-related content into instruction seems to be an open question. Research on teachers' professional knowledge about teaching technology-related content, in general, does not refer to the TPACK framework, but introduces new constructs. Examples are "Teachers' emphasis on developing students' digital information and communication skills" (Siddiq et al., 2016) and "Teaching in a Digital Environment Capacity" (Claro et al., 2018). In terms of theory building, it may be promising to investigate how these currently separated streams of research may inform each other in order to develop a parsimonious framework that covers both how to teach and what to teach in the context of digital transformation. In this vein, Graham (2011) pointed to an important aspect: the value of such a framework depends on its ability to predict meaningful outcomes. In the end, the desired outcome would be gains in student learning. However, the effects of the professional knowledge of an individual teacher on student learning may be hard to evaluate due to many confounding variables (Darling-Hammond, Newton, & Wei, 2010). Following the theory of planned behavior and technology acceptance models (Ajzen, 1991; Davis, 1989; Venkatesh, Thong, & Xu, 2012), a meaningful proximal outcome could be the intention to use or the actual use. These models imply that beyond professional knowledge further predictors may be considered. Following Teo and van Schaik (2012) and Scherer, Siddiq, and Tondeur (2019), attitudes, in particular, play an important role in predicting teachers' use of technology.

The aim of the study at hand is to develop and to validate a framework of in-service teachers' technology-related knowledge and attitudes in order to predict the use of technology as a means of instruction and as a content of instruction (Use). The purpose of this prediction is to obtain a better understanding for the professional development of teachers' technology-related knowledge and attitudes. Hence, we do not consider factors beyond the control of teachers, such as school characteristics, e.g., school equipment. This may be an issue because school equipment and resources, e.g., digital devices and high-speed internet, seem to play an important role for Use (European Commission, 2019; Fraillon et al., 2019; in contrast Gil-Flores, Rodríguez-Santero, & Torres-Gordillo, 2017). Since we do not consider school level variables (e.g., school equipment), although they might be important for Use, we will come back to this point in the method section. Overall, our research questions are:

RQ1. What technology-related knowledge and attitudes predict in-service teachers' use of technology as a means of instruction?

RQ2. What technology-related knowledge and attitudes predict in-service teachers' use of technology-related content in instruction?

In answering the two research questions, we aim at combining the two streams (how and what to teach) in order to identify technology-related professional knowledge and attitudes that are necessary for teaching with and about technology. By doing so, we could inform professional development programs of in-service teachers that aim at fostering technology-related knowledge and attitudes.

2. Theoretical framework

Adopting technology for learning purposes is a complex endeavor (Straub, 2009). Two kinds of models may be suitable for explaining teachers' use of technology: (1) technology acceptance models, and (2) the will, skill, tool (WST) model (Christensen & Knezek, 2008).

(1) With respect to teachers' acceptance and use of technology, Scherer, Siddiq, and Tondeur (2020) carried out a meta-analytical structural equation modeling study. They found that technology acceptance may only directly predict the use of technology instead of being mediated by behavioral intention. Although technology acceptance models are developed for predicting the use of technology, the underlying concept may also be helpful for predicting the integration of technology-related content into instruction. Technology acceptance models draw from the theory of planned behavior, which aims at explaining behavior in general (Ajzen, 1991). In the context at hand, it would imply that attitudes, subjective norms, and perceived behavioral control predict the intention to use technology-related content in instruction. Following Scherer et al. (2020), they may directly predict the actual use instead of behavioral intention. (2) The WST model has successfully been used to predict teachers' technology use (Agyei & Voogt, 2011; Petko, 2012). 'Will' can be defined as "a positive attitude toward the use of technology in instruction", 'skill' as "the ability to use and experience technology", and 'tool' as "availability, accessibility and extent of use of technology" (Knezek & Christensen, 2016, p. 311). Since from a professional development perspective, the tool component of the WST model may be without the control of individual teachers, we do not consider it in our framework. This is also the case for the social norm component of the theory of planned behavior, i.e., attitudes and behavioral control remain as predictors. From the individual teachers' perspective, we hypothesize their will (attitudes) and skill (behavioral control) to predict Use. Due to the purpose of our research, we restrict ourselves to technology-related knowledge and attitudes, i.e., we do not consider pedagogical content knowledge of specific subjects.

Drawing from the theory of planned behavior, we define attitudes as "the degree to which a person has a favorable or unfavorable evaluation or appraisal of the behavior in question" (Ajzen, 1991, p. 188) and hypothesize:

H1. Positive attitudes towards technology in instruction positively predict the use of technology in instruction.

H2. Positive attitudes towards technology as content positively predict the use of technology-related content.

Petko (2012) suggested utilizing the TPACK framework in order to specify the skill component of the WST model. When it comes to the use of technology as a means of instruction across subjects, technological pedagogical knowledge (TPK) is the pertinent construct (Zhang, Liu, & Cai, 2019). In comparison to TPACK, it addresses knowledge relevant for all kind of subjects. For instance, cognitive activation using technology may play an important role in all kind of subjects. However, its manifestation in specific subjects could vary. Koehler, Mishra, and Cain (2013, p. 16) define TPK as an "understanding of how teaching and learning can change when particular technologies are used in particular ways". To specify this construct, it may be helpful to utilized research about PK. Voss, Kunter, and Baumert (2011) identified 'teaching methods', 'students' heterogeneity', 'classroom assessment', and 'classroom management' as facets of PK. TPK could be conceptualized as knowledge about how technology facilitates or changes those facets, e.g., how classroom management works in a technology-rich environment. We hypothesize:

H3. TPK positively predicts the use of technology in instruction.

When teaching about technology using technology, technology is the content as well as the means of instruction. In a formal sense, the letter 'C' of TPACK could be replaced by 'T'. Teaching with technology about technology may be a special form of TPACK that Koehler et al. (2013, p. 16) define as "an understanding that emerges from interactions among content, pedagogy, and technology knowledge". We refer to the DigComp 2.1 framework (Carretero et al., 2017) to delineate the content (= technology). This is consistent with the DigCompEdu framework (Redecker, 2017). Relying on DigComp may have two advantages. First, instructional content is normative in nature and usually codified in official curricula or standards. Referring to the framework of the European Union may be a common factor across (European) countries. Second, the DigComp framework is, as the current version 2.1 implies, dynamic. By referring to DigComp, our own framework is also dynamic and keeps track of technological developments. Concerning this specific TPACK, i.e., knowledge about teaching with technology about technology, we hypothesize:

H4. TPACK positively predicts the use of technology-related content.

TPK and TPACK are associated (Koehler et al., 2013), the latter referring to a specific subject; hence, cross-disciplinary TPK may be conducive to TPACK. A teacher who has more knowledge about instructing with technology across various subjects may also demonstrate increased knowledge about teaching a specific subject, e.g., technology-related content. Indeed, Koh, Chai, and Tsai (2013) as well as Schmid, Brianza, and Petko (2020) provided empirical evidence for a positive influence of TPK on TPACK. We hypothesize:

H5. TPK positively predicts TPACK.

As Koehler et al. (2013) posit, TK is hard to define and if referring to specific technology would quickly lead to an outdated definition. As outlined for H4, we rely on the five facets of the DigComp 2.1 framework to delineate 'technology', i.e., teachers are expected to show the same kind of knowledge that they are supposed to foster among their students. TK may be the prerequisite for all other technology-related professional knowledge (Schmidt et al., 2009). In order to judge how technology can support the pursuit of pedagogical goals (in a specific subject), TK is necessary. In our context, TK also represents CK. Hence, just like CK, in general, we expect TK to positively predict TPACK. Indeed, Koh et al. (2013) showed that TK positively predicts both TPK and TPACK. We hypothesize:

H6. TK positively predicts TPK.

H7. TK positively predicts TPACK.

Professional learning communities or communities of practice (teacher communities) play an important role in the professional development process of teachers (Borko, 2004; Prenger, Poortman, & Handelzalts, 2019; Vangrieken, Meredith, Packer, & Kyndt, 2017; Warwas & Helm, 2018). This may also be the case in the realm of technology integration (Lawless & Pellegrino, 2007; Tseng & Kuo, 2014; Zhang et al., 2019). The activities of teacher communities may include the development of school curricula (Supovitz, 2002) or knowledge exchange (Vangrieken et al., 2017). Vangrieken, Dochy, Raes, and Kyndt (2015) summarize such activities as teacher collaboration. Tondeur, van Braak, Siddiq, and Scherer (2016) integrate 'collaboration' into their model for the preparation process of pre-service teachers for using technology in instruction. Agyei and Voogt (2012) report a positive influence of collaboration on the development of various facets of the TPACK framework among pre-service teachers. From a professional development perspective, the necessary knowledge to participate in such teacher communities may be put into focus. Since the study at hand restricts itself to technology, our construct Technological Collaboration Knowledge (TCoK) addresses teachers' knowledge about how to use technology in order to collaborate with other teachers, e.g., using databases to exchange knowledge. Drawing from the theory of planned behavior, such knowledge may be an important (indirect) predictor for actual collaboration: Teachers with higher TCoK may collaborate more in terms of quality and quantity, which positively predicts TPK and TPACK. We hypothesize:

H8. TCoK positively predicts TPK.

H9. TCoK positively predicts TPACK.

Since TCoK has, as the name implies, a technology component, we expect TK to be conducive for TCoK. Teachers who have general knowledge about digital content creation and collaboration using technology may also be able to use this general knowledge to collaborate with their colleagues. We hypothesize:

H10. TK positively predicts TCoK.

Fig. 1 summarizes our conceptual framework.

3. Method

3.1. Instruments

The validity of a measurement depends on its intended use (AERA, APA, & NCME, 2014). Self-assessment instruments may be suitable from a professional development perspective. As the theory of planned behavior implies, self-perceptions (indirectly) predict behavior. Self-assessment instruments tend to capture self-efficacy, which could be a good predictor for teachers' use of technology (Scherer, Tondeur, & Siddiq, 2017; Voogt et al., 2013). Moreover, self-assessment instruments are well established in research about technology acceptance (Nistor, 2014), the WST model (Knezek & Christensen, 2016; Petko, 2012), and TPACK (Scherer et al., 2017; Schmid et al., 2020; Valtonen et al., 2017). Against this backdrop, we utilize a self-assessment instrument to capture the constructs of our conceptual framework. The development of the instrument was informed by fourteen expert interviews. These experts show a diverse background: teachers, school principals, educational policy makers, federation representatives, and researchers in the realm of digital transformation. A first draft of the instrument was pretested with a sample of 132 teachers and, based on this, refined. By means of focus group discussions with the school management teams at five partner schools, we compiled the final instrument (Seufert, Guggemos, Tarantini, & Schumann, 2019). For the purpose of this study, we reduced the complexity of the DigComp 2.1 related measurement instruments, namely TK and TPACK. In the original version they have a higher-order structure where the five facets of the DigComp 2.1 framework are captured with several items and these five facets operationalize TK and TPACK. For the purpose of parsimony, we measure TK and TPACK with five items corresponding to the five facets of the DigComp 2.1 framework. Convergent validity may be achieved because the latent correlation of the constructs measured with the full instruments and the reduced ones utilized in the present study are greater than 0.912. All the used items in this study can be found in S1 in the Appendix. The knowledge and attitude constructs are captured using a seven-point scale of rating ranging from "does not apply at all" to "applies very strongly". A five-point scale of rating is utilized to measure the items for Use: never, infrequently (1–2 times per semester), occasionally (3–5 times per semester), frequently (every month), and very frequently (every week).

3.2. Data collection and sample

Our sample comprises 212 in-service teachers at commercial vocational schools in German-speaking Switzerland. To obtain the sample, we contacted all commercial vocational schools in German-speaking Switzerland. Besides the five partner schools, which were involved in the development of the instrument, four further schools participated. An online questionnaire was given to all teachers at the nine schools. Participation was voluntary and anonymous. We checked for potential outliers by means of univariate (boxplots) and multivariate (Mahalanobis distances) methods. We inspected all of the potential five outliers and decided not to exclude any observations. Table 1 summarizes the key characteristics of our sample and any missing values in the context variables. For all other variables (see S1 in the Appendix), 3.2% of missing values occurred. Due to a non-significant Little's MCAR test (χ^2 (1612) = 1834.321, p = 1) and having inspected the missing data patterns, we treat them as missing completely at random.

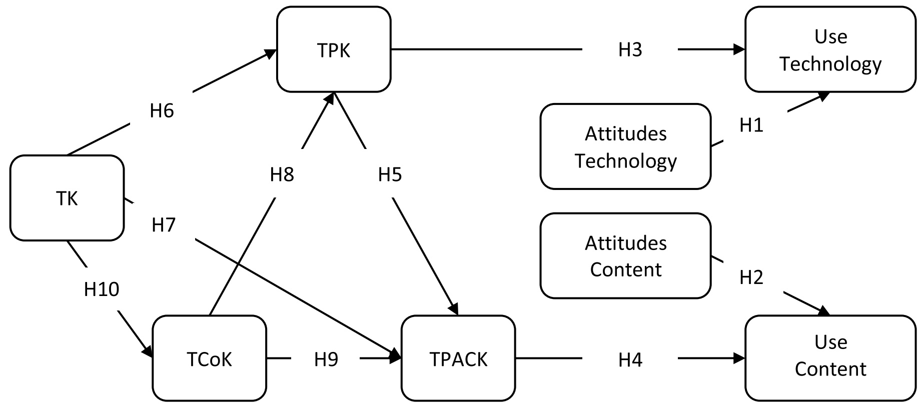


Fig. 1. Conceptual framework for predicting use of technology as a means of instruction and use of technology as a content of instruction with teachers' professional knowledge and attitudes.

Table 1
Characteristics of the sample (N = 212).

Variable	Manifestation	Frequency	Percent
Gender	Female	105	49.5
	Male	105	49.5
	Missing	2	1.0
Age [years]	<36	22	10.4
	36–45	47	22.2
	46–55	87	41.0
	56–65	56	26.4
	Missing	0	0.0
Teaching experience [years]	1–5	25	11.8
	6–10	36	17.0
	11–15	36	17.0
	16–20	35	16.5
	>20	77	36.3
	Missing	3	1.4

3.3. Data analysis

First, we group mean center the covariates to control for heterogeneity between the nine schools in our sample (Bell, Jones, & Fairbrother, 2018). Teachers' professional knowledge and attitudes are contrasted with the knowledge and attitudes of other teachers from the same school (group mean), instead of from the entire sample (grand mean). By removing the between-school variance, the estimated coefficients only represent individual level effects. This tackles the issue that we do not consider the tool component of the WST model; the reported associations are not influenced by varying equipment, resources, and support across the nine schools.

Since we aim at testing the structure of our conceptual framework, we rely on covariance-based structural equation modeling (CB-SEM) (Hair, Risher, Sarstedt, & Ringle, 2019). Our data, measured, at the least, on five-point scales of rating, show slight deviations from a normal distribution. Moreover, teachers are nested in nine schools. Hence, we utilize a maximum likelihood estimator (MLR) that is robust against slight deviations from a normal distribution and non-independence of observations (cluster robust standard errors). In light of this, we report χ^2 statistics after considering the Satorra Bentler correction: SB-χ^2 (Satorra & Bentler, 2010). Due to our assumption of data missing completely at random, we rely on full information maximum likelihood to handle the missing data (Jia & Wu, 2019). We use the R-package lavaan 0.6–5 for all CB-SEM based analyses (Rosseel, 2012).

For assessing the quality of our measurement instrument, Cronbach's α and McDonalds' ω act as measures for reliability (Scherer et al., 2017). A cut-off value for an acceptable reliability could be 0.7 (Hair et al., 2019). Convergent validity of the measures may be ensured if the average variance extracted (AVE) is larger than 0.5; discriminant validity if the heterotrait–monotrait (HTMT) ratio is smaller than 0.9 for conceptual similar constructs (Hair et al., 2019). The HTMT ratio is defined as the mean value of the item correlations across constructs relative to the geometric mean of the average correlations for the items measuring each construct. The HTMT ratio has been demonstrated to be superior to the Fornell-Larcker criterion for detecting lack of discriminant validity (Franke & Sarstedt, 2019). A rigorous assessment of discriminant validity may be important because Scherer et al. (2017) showed substantial similarity between the facets of the TPACK framework. To assess the overall model fit of the framework, we use CFI and TLI as comparative measures, and RMSEA and SRMR as absolute ones. The following values serve as cut-off values (Van de Schoot, Lugtig, & Hox, 2012): acceptable fit: CFI and TLI > 0.90, RMSEA < 0.08, SRMR < 0.10; good fit: CFI and TLI > 0.95, RMSEA < 0.05, SRMR < 0.06. In case of an acceptable model fit, we modify our measurement model (Schmid et al., 2020) in order to achieve a good fit.

As can be seen from Fig. 1, our conceptual framework contains indirect paths that imply mediation, e.g., from TK via TPK to Use Technology. For the theoretical soundness of the conceptual framework, it is important to evaluate the kind of mediation involved. To this end, we rely on Zhao, Lynch, and Chen (2010) who revised the approach of Baron and Kenny (1986). Following Zhao et al. (2010), an indirect-only mediation could indicate that the mediators in the model are consistent with the theoretical framework. It is characterized by a significant indirect path and, at the same time, non-significant direct path. In our case, for instance, this would imply that the indirect path from TK to Use Technology is significant, but a direct path from TK to Use Technology is not significant. If the direct path was also significant, this would point to an incomplete theoretical framework: there may be further not-yet-considered mediators hidden in the 'direct' path. To assess the statistical significance of the indirect paths, we use bias-corrected bootstrapping with 5000 samples to create the 95% confidence intervals (Preacher & Hayes, 2008). If the 95% confidence interval does not include zero, the indirect path is regarded as significant.

Observed and unobserved heterogeneity may influence the reported path coefficients. In terms of observed heterogeneity, moderators could be teachers' gender, age, and main subject (Siddiq et al., 2016). Due to convergence problems, we cannot rely on CB-SEM to check for moderators by means of multigroup analyses. The reason is the relatively small sample size in our study. As an alternative, we rely on partial least squares SEM (PLS-SEM) that has, in comparison to CB-SEM, smaller formal sample size requirements (Hair et al., 2019). In the case of our conceptual framework, fifty teachers in each group could be sufficient. Before carrying out multigroup analyses, we check for measurement invariance by means of permutation tests (Henseler, Ringle, & Sarstedt, 2016). In terms of age, we form three groups: under 46 years (junior, n = 69), 46–55 years (middle, n = 87), and 56+ years (senior, n = 56).

Concerning the main subject, there are only two groups with a sufficient sample size: business education (n = 70) and language teachers (n = 66). Finally, we use PLS-SEM to control for unobserved heterogeneity by means of a finite mixture segmentation. Evidence for the absence of unobserved heterogeneity may be a lower information criterion AIC3 and CAIC for a one-class model in comparison to every multiclass solution (Sarstedt, Becker, Ringle, & Schwaiger, 2011). We utilize SMART-PLS 3.3.2 for all PLS-SEM related analyses.

4. Results

All response patterns to the items (see S1 in the Appendix) show significant deviations from a normal distribution (Shapiro Wilk test: $p < .05$). However, the deviations are not severe. The absolute values for skewness and kurtosis are below 1.14 and 4.99, respectively. The teachers in the sample show an overall positive attitude toward technology as a means of instruction and technology-related content, which is indicated by a mean on the seven-point scale of rating of 5.45 and 4.99, respectively. A mean of 2.308 and 2.357 on a five-point scale of rating for Use Technology and Use Content, respectively, points to an infrequent to occasional use.

The quality of the measurement model is decent (see Table 2). Cronbach's α and McDonalds' ω is greater than 0.7 for all constructs. The AVE is greater than 0.5 with one exception. For TK it equals 0.475, which is only slightly below the cut-off value of 0.5. The reason is item 12 (see S1 in the Appendix) that shows a factor loading of 0.59. Overall, however, convergent validity may be achieved. Discriminant validity might also be ensured. The highest HTMT ratio appears for TPK and TPACK with 0.813. This is well below the cut-off value of 0.9. Even the 95% confidence interval of this HTMT ratio is below 0.90 [0.730, 0.881].

The initial structural equation model shows an acceptable fit SB-χ^2 (362) = 565.229 ($p < .001$), CFI = 0.942, TLI = 0.935, RMSEA = 0.051 (90% CI = [0.044, 0.059]), SRMR = 0.060. By including four residual correlations between items, we obtained a good fit. Fig. 2 shows these residual correlations and the fit values of the adjusted model. The considered residual correlations may be reasonable from a conceptual point of view. For instance, TK items 12 and 13 (see S1 in the Appendix) address rather technical aspects whereas the other TK items cover conceptual aspects like information literacy and communication. However, there are no differences between the initial structural model and the adjusted model (with residual correlations) in terms of sign and statistical significance. All hypothesized paths are significant at the 5% level. The evidence reported throughout this paper is based on the adjusted model.

As Table 3 shows, all the indirect paths in our conceptual framework are statistically significant. To assess the type of mediation, we inserted (at once) potential direct paths. However, they are all non-significant: TK → Use Technology (−0.054, p = .828), TK → Use Content (−0.178, p = .268), TCoK → Use Technology (−0.062, p = .772), TCoK → Use Content (−0.020, p = .892), TPK → Use Content (−0.044, p = .804). These findings of significant indirect paths and, at the same time, non-significant direct paths point to indirect-only mediations.

In terms of the multilevel structure (teachers nested in nine schools), the intraclass correlation coefficient ICC(1) of the two outcome variables Use Technology and Use Content equals 0.149 and 0.001, respectively. For all other variables, the ICC(1) ranges between 0.004 (TK) and 0.065 (TPK). Although the between-school variance may not be small enough to be ignored, group mean centering does not have any influence on the signs of the path coefficients and their statistical significance at the 5% level. Consistent with the substantial between school variance for Use Technology, the explained variance drops from 45.9% to 36.0% when using group mean centered predictors instead of predictors derived from the raw data.

The PLS-SEM measurement model and structural model showed good properties (Hair et al., 2019). Moreover, PLS-SEM yielded the same results as CB-SEM in terms of sign and significance at the 5% level. The results of the PLS-SEM multigroup analyses are depicted in Table 4. There are differences in path coefficients for all three hypothesized moderators. Worth noting, although there are no differences in path coefficients, compositional measurement invariance seems to be an issue between junior teachers, on the one hand, and middle as well as senior teachers, on the other. The problematic items are 15 (junior vs. senior), and 17 and 18 (junior vs. middle), see S1 in the Appendix for items. All these items pertain to TPK.

Concerning unobserved heterogeneity, the finite mixture segmentation indicated a one-class solution as the best trade-off between model complexity and parsimony: one class: AIC3 = 2,509, CAIC = 2,559; two classes: AIC3 = 2,519, CAIC = 2,623; three classes: AIC3 = 2,532, CAIC = 2,690. All models with more than three classes also show worse AIC3 and CAIC in comparison to the one-class solution.

Since our conceptual framework (see Fig. 1) is not the only possible one, we compared it with competing models by including reasonable further paths. TCoK could positively predict TK. Moreover, TPK and TPACK might predict TCoK. Besides this, it is an open question if the specific TPACK that refers to the instruction of technology related content could also positively predict Use Technology that addresses teaching with technology across subjects (Petko, 2020). Moreover, TPK might directly predict Use Content besides the already considered path via TPACK. However, none of these paths is statistically significant at the 5% level: TCoK → TK (0.207, p = .443), TPK → TCoK (0.187, p = .293), TPACK → TCoK (0.152, p = .484), TPK → Use Content (−0.046, p = .706), TPACK → Use Technology (−0.012, p = .924).

5. Discussion and limitations

The aim of our research was to predict teachers' use of technology as a means of instruction (RQ1) and the use of technology-related content in instruction (RQ2) for the purpose of professional development. The findings lend support to our conceptual framework, depicted in Fig. 1, because all hypotheses were accepted. Moreover, CB-SEM and PLS-SEM

Table 2
Descriptives for constructs, assessment of measurement model, and latent correlations among constructs (N = 212).

–						Latent correlation among constructs							
	Construct	α	ω	AVE	HTMT	1	2	3	4	5	6	7	8
1	Attitudes Technology	.883	.897	.804	<.698	1.00							
2	Attitudes Content	.924	.925	.861	<.698	.669	1.00						
3	TCoK	.874	.904	.664	<.685	.541	.516	1.00					
4	TK	.853	.773	.475	<.722	.644	.639	.656	1.00				
5	TPK	.915	.916	.685	<.813	.598	.586	.627	.634	1.00			
6	TPACK	.903	.906	.652	<.813	.614	.650	.692	.692	.821	1.00		
7	Use Technology	.883	.891	.621	<.646	.523	.347	.357	.410	.571	.495	1.00	
8	Use Content	.805	.811	.592	<.670	.481	.507	.442	.442	.553	.669	.610	1.00

Note. α = Cronbach's α, ω = McDonald's ω, AVE = Average variance extracted, HTMT = Heterotrait–monotrait ratio. All correlation significant at the 0.1% level (two-sided) considering the cluster structure of the data (teachers nested in nine schools). Constructs 1–6 are measured on a seven-point scale of rating, constructs 7 and 8 on a five-point scale of rating. Item loadings, means, and standard deviations are reported in S1 in the Appendix.

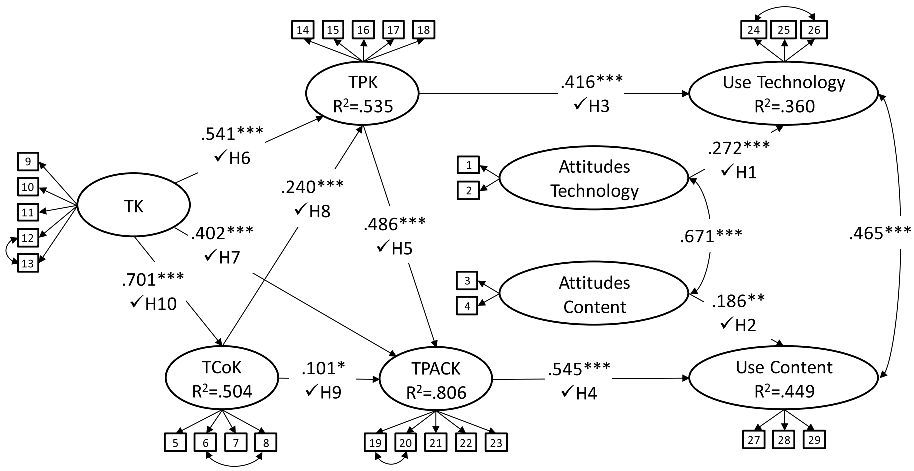


Fig. 2. Structural equation model to predict 'Use Technology' and 'Use Content' (N = 212). Note. Standardized path coefficients. Assessment of significance based on cluster robust standard errors (cluster = schools). *p < .05, **p < .01, ***p < .001. Fit values: SB-χ^2 (358) = 513.121 (p < .001), CFI = 0.956, TLI = 0.950, RMSEA = 0.048 (90% CI = [0.038, 0.057]), SRMR = 0.050. For reasons of clarity, the latent correlation between Attitudes Technology and Attitudes Content with TK is not depicted (0.699 and 0.705, p < .001). All factor loadings, see S1 in the Appendix, are significant (p < .001).

Table 3
Indirect paths of the conceptual framework (N = 212).

			Specific indirect			Total indirect		
Outcome	Mediator(s)	Predictor	Std. Coef.	Coef.	95% CI	Std. Coef.	Coef.	95% CI
Use Technology	TPK	TK	0.225	0.200	(0.081, 0.319)	0.296	0.263	(0.129, 0.396)
	TCoK, TPK		0.071	0.063	(0.033, 0.093)			
	TPK	TCoK	0.100	0.070	(0.003, 0.215)	0.100	0.070	(0.033, 0.107)
Use Content	TPACK	TK	0.219	0.177	(0.091, 0.263)	0.447	0.361	(0.283, 0.439)
	TCoK, TPK, TPACK		0.045	0.037	(0.012, 0.061)			
	TCoK, TPACK		0.039	0.032	(0.003, 0.060)			
	TPK, TPACK		0.144	0.116	(0.039, 0.193)			
	TPACK	TPK	0.265	0.196	(0.099, 0.292)	0.265	0.196	(0.099, 0.292)
	TPACK	TCoK	0.055	0.035	(0.001, 0.070)	0.119	0.076	(0.018, 0.134)
	TPK, TPACK		0.064	0.041	(0.009, 0.073)			

Note. Coef. = unstandardized path coefficients. Std. Coef. = standardized path coefficients. Confidence intervals based on bias corrected bootstrapping with 5000 samples.

Table 4
Multigroup analyses: Significant different path coefficients.

Moderator	Manifestation	Path	Std. Coef.	|Dif.|	p-value
Gender	Male (n = 105)	TK → TCoK	0.408	0.265	.006
	Female (n = 105)		0.673		
	Male (n = 105)	TPACK → Use Content	0.377	0.228	.043
	Female (n = 105)		0.605		
Age	middle (n = 87)	TK → TPACK	0.391	0.263	.025
	senior (n = 56)		0.128		
	middle (n = 87)	TPK → TPACK	0.318	0.291	.027
	senior (n = 56)		0.609		
Main subject	Business (n = 70)	TCoK → TPK	0.134	0.445	.010
	Language (n = 66)		0.578		
	Business (n = 70)	TK → TPK	0.569	0.516	.003
	Language (n = 66)		0.053		

Note. Std. Coef. = standardized path coefficients, Dif. = difference in path coefficients. P-values based on Bootstrappping (BCa) with 5000 samples. Compositional measurement invariance is ensured (Henseler et al., 2016).

yielded identical results in terms of sign and statistical significance at the 5% level. This multimethod approach provides evidence for the robustness of our findings (Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016). Furthermore, it may, indeed, be necessary to separate the constructs Use Technology and Use Content. They can clearly be separated (HTMT ratio = 0.646) and their latent correlation of 0.465 is significantly below 1 (see Fig. 2). The correlations among TK, TPK, and TPACK found in our study are well in line with other studies based on self-assessment instruments (Koh et al., 2013; Schmidt et al., 2009). For instance, the manifest correlation between TPK and TPACK equals .74, exactly the same value reported by Koh et al. (2013). In comparison to those studies, however, the correlation in our study is comparatively higher between TK and TPACK. This is in line with the conception that TK also represents CK when teaching with technology about technology. Overall, the measurement instrument shows decent reliability ($\alpha > 0.853$), convergent validity (AVE > 0.475), and discriminant validity (HTMT ratio < 0.813). The overall fit of the model, when considering four residual correlations between items, is good: CFI = 0.956, TLI = 0.950, RMSEA = 0.048, SRMR = 0.050. To check for substantially relevant local model misspecifications, we also inspected the expected parameter change in combination with the modification index (Saris, Satorra, & van der Veld, 2009); however, we did not detect any signs of this. As Scherer et al. (2017) showed, it may be hard to separate the constructs of the TPACK framework. Our measurement instrument might ensure, at the very least, an acceptable separation of the constructs (discriminant validity). In terms of the quality of our prediction, we were able to explain 36.0% and 44.9% of the variance for Use Technology and Use Content, respectively. This is in line with Agyei and Voogt (2011) who reported an R^2 when considering 'will' and 'skill' of 41%–42% (depending on the model) for integration of technology. Knezek and Christensen (2016) claim explained variances of >60% for the WST model. However, a fair comparison is not possible because in our research we only consider will and skill as we focused on the individual teacher's technology-related knowledge and attitudes.

To judge the theoretical soundness of our conceptual framework, it was important to determine the kind of mediation. Our findings of indirect-only mediations may be evidence for a correctly specified conceptual framework (Zhao et al., 2010). Namely, the direct paths from TK to Use are non-significant. It indicates that only TPK and TPACK directly predict Use. Moreover, TPK does not directly predict Use Content but the indirect path via TPACK is significant. This finding may indicate that considering the domain-specific TPACK in addition to the domain-independent TPK yields additional insights and should be retained. Furthermore, TCoK is not directly associated with Use. Rather, it predicts TPK and TPACK and, indirectly, Use. This is in line with the conceptualization of the construct as an important aspect in the professional development process of teachers in the context of digital transformation. Again, CB-SEM and PLS-SEM led to identical results for all these analyses.

The multigroup analyses in combination with measurement invariance analyses yielded some remarkable findings. First, the latter analyses indicated a substantially different understanding of TPK between junior teachers (<46 years old), on the one hand, and middle and senior teachers (46+ years old) on the other. This may not be surprising because pedagogical approaches and technology have substantially changed over time. Our findings may also caution against taking measurement invariance in terms of age for granted. Apart from age, we found evidence for compositional measurement invariance in terms of gender and main subject. This might point to a fairly similar understanding across gender and teachers with different main subjects. What is remarkable is the substantially higher association between TK and TPK for business education teachers in comparison to language teachers (0.569 vs. 0.053). Besides this, however, the path coefficients in the structural model are fairly similar. In summary, the main source of observable heterogeneity between teachers might be age. Teachers in the middle and senior group (46+ years old) seem to have a different understanding of TPK in comparison to teachers who are younger. These findings may be in line with the ICILS (Fraillon et al., 2019). Despite some differences in path coefficients, however, the result of a one-class solution in the finite mixture segmentation points to a one-class solution for the entire sample as a good trade-off between model complexity and parsimony. Overall, we regard our conceptual framework as robust against various forms of heterogeneity.

Our study has several limitations that have to be mentioned. Due to cross-sectional data, we cannot make any causal claims. Longitudinal data would be more suitable to investigate the relationships between the constructs in our study, e.g., by using cross-lagged panel designs. Moreover, our sample is limited. It comprises only 212 in-service teachers from nine commercial vocational schools in German-speaking Switzerland who voluntarily participated in the study. At least the sample size may not be an issue. As a robustness check, we carried out factor score regression. In case of small samples sizes, using regression factor scores could yield higher precision in comparison to structural equation modeling (Loncke et al., 2018). To this end, we utilized the syntax provided by Loncke et al. (2018). The fit of the factor score model is good: SB-χ^2 (14) = 19.235 (p = .156), CFI = 0.995, TLI = 0.992, RMSEA = 0.043 (90% CI = [0.000, 0.087]), SRMR = 0.028. The path coefficient of the factor score regression and those of the structural equation model are identical in terms of sign and statistical significance at the 5% level. The second-level sample size of only nine schools prevented us from investigating school level relationships, i.e., to take the multilevel structure fully into account. Furthermore, we relied solely on teachers' self-reports, which may raise concerns about common methods bias (Conway & Lance, 2010). For Use, in particular, it would be beneficial to have objective measures. Moreover, it could be revealing to take, besides the quantity of Use, also its quality into account (Sailer et al., 2021; this issue). In this vein, however, it has to be stressed that any measurement can only be evaluated with respect to its purpose (AERA et al., 2014). We agree with Backfisch, Lachner, Hische, Loose, and Scheiter (2020) that the self-assessment instruments of TPK and TPACK have disadvantages. If the purpose is to capture professional knowledge in order to support professional development, however, it should be taken into account that performance tests in high-stakes contexts could have negative side effects, e.g., evoke negative emotions among teachers (Ryan & Weinstein, 2009).

Moreover, we did not consider the actual degree of collaboration among teacher. The direct association of TCoK and TPK as well as of TCoK and TPACK (see Fig. 1) is likely be mediated by the actual level of collaboration. In further studies, it may be revealing to include the level of technology-supported collaboration (quality and quantity) as a mediator.

6. Implications

6.1. Theoretical implications

The paper at hand contributes to the knowledge base about the TPACK framework in many ways. Graham (2011) pointed out three problems in TPACK research: 1) What, 2) How, and 3) Why. Our research also addresses all three areas.

1) What. The trade-off between comprehensiveness and parsimony guided the development of our conceptual framework for predicting the use of technology as a means and as a content of instruction. Currently, the two types of use are discussed separately. We argue that when teaching with technology about technology, technology also represents the content. The corresponding knowledge may be regarded as TPACK. Based on this assumption, it is not necessary to integrate a specific construct for teachers' knowledge on teaching with technology, about technology, into the conceptual framework. Moreover, this assumption reduces the complexity of the TPACK framework because CK and its interactions with other constructs, e.g., TCK, equal TK.

Since the publication of the well-established TPACK self-assessment instrument of Schmidt et al. (2009), research on teachers' professional knowledge has progressed. Hence, we took into account recent research about teachers' general pedagogical knowledge (Depaepe & König, 2018; Voss et al., 2011) and high-quality instructional settings (Praetorius, Pauli, Reusser, Rakoczy, & Klieme, 2014). Referring to research about PK could contribute to the theoretical founding of TPK and keep such knowledge up to date as regards the state of research about high-quality learning settings.

According to Graham (2011), as well as Koehler and Mishra (2009), it is especially difficult to capture the technology facet. We relied on the DigComp 2.1 framework (Carretero et al., 2017), which may have two advantages. DigComp 2.1 is, as the name implies, dynamic. Hence, new technological developments are automatically integrated into our conceptual framework too. Moreover, it contributes to the parsimony of our conceptual framework: the DigComp framework is used to delineate TPACK, as well as TK.

TPACK and TK are normative in nature. In our case, the DigComp framework determines what knowledge teachers are expected to foster among students (TPACK) and should possess themselves (TK). This is legitimate because we developed the conceptual framework in collaboration with the schools in our sample; the school management teams agree with the framework. However, researchers or practitioners who want to adapt our instrument have to examine whether the conception of TK and TPACK is in line with curricula or standards pertinent to them. Moreover, the instruction should also be considered (Pellegrino, DiBello, & Goldman, 2016), i.e., the measurement instrument should be consistent with the content that teachers are taught or are supposed to develop informally.

Although the importance of collaboration in the professional development of teachers is well documented, this aspect has rarely been explicitly considered in quantitative research about teachers' technology integration. The consideration of the construct TCoK in the conceptual framework was highly recommended by the school representatives involved in the development process (Seufert, Guggemos, Tarantini, & Schumann, 2019). From their point of view, professional development in the context of digital transformation requires joint effort and, thus, TCoK. Our findings lend support to this assertion as we show that at the individual teacher level, TCoK significantly predicts TPK and TPACK and, indirectly, Use. However, further research is necessary for a better understanding of the role of TCoK in the professional development process. The actual level of collaboration and its quality are likely to mediate the relationship between TCoK and TPK and TPACK. Moreover, in line with the theory of planned behavior, attitudes towards technology-related collaboration should be considered.

2) How. Angeli and Valanides (2009) criticized the unclear relationship among the constructs of the TPACK framework. Koh et al. (2013) posited a relationship among the constructs of the TPACK framework and provided evidence for their hypotheses. We were able to replicate the model structure of Koh et al. (2013). The added value of our work when compared to Koh et al. (2013) is the mediation analysis. As we showed, not all constructs, e.g., TK, directly predict Use.

Our findings may contribute to a better understanding of the relationships between the constructs of the TPACK framework from a prediction perspective. For a factor analytical perspective, see Scherer et al. (2017).

The delineation of the constructs in the TPACK framework is a regular problem. As we showed, our measurement model has sufficient discriminant validity (HTMT ratio < 0.813). Moreover, the overall model fit indices are good (CFI $= 0.956$, SRMR $= 0.050$). By developing a refined TPACK measurement instrument, we might have followed Scherer et al. (2017, p. 13) who "encourage the further development of TPACK measures that might distinguish more clearly between the different knowledge dimensions and further validation studies that investigate the substantial meaning of the specific TK factor".

3) Why. TPACK research rarely addresses the usefulness of this framework to predict meaningful outcomes. An exception for in-service teachers is Hsu (2016) who integrated the TPACK framework into a technology acceptance model to predict the adoption of mobile-assisted language learning. We combined the TPACK framework with the WST model to predict teachers' use of technology as a means of instruction and as a content of instruction. Both forms of use may be important in the context of digital transformation (Fraillon et al., 2019). In line with the WST model and the theory of planned behavior, our findings show that besides the TPACK constructs, attitudes should be considered for predicting Use. Solely relying on the TPACK framework may not be sufficient if the aim is to predict Use (Petko, 2012). However, TPK and TPACK might be the most important predictors. In comparison to TPK and TPACK (skill), attitudes (will) show a lower predictive power. This is in line with Agyei and Voogt (2011) who reported skill as the most important predictor among pre- and in-service teachers. Interestingly, for pre-service teachers' technology integration, will seems to be a better predictor than skill (Farjon, Smits, & Voogt, 2019). Such differences between pre- and in-service teachers could be promising avenues for further research.

6.2. Practical implications

As we have shown, TK has no direct effect on Use. TK in isolation may neither be sufficient to teach with technology nor about technology. Thus, in teacher education and training TPK and TPACK may be brought into focus. However, TK seems to be conducive to TPK, TPACK, and TCoK and, therefore, should not be neglected. Formative assessments could help to identify teachers with insufficient TK. Such teachers might receive offers for specific training. Overall, it could be advantageous to rely on an overarching framework for teacher and student TK that is accepted by the stakeholders, e.g., the DigComp framework. By means of this, teachers are trained in the same knowledge that they are also expected to foster among their students, e.g., efficient technology-supported collaboration, and complexity can be reduced. From the professional knowledge facets, only TPACK directly predicts the use of technology-related content. Hence, using technology as a content of instruction may require a specific kind of knowledge. Efforts towards TPACK development may build upon TK and TPK, i.e., TPACK is addressed after TPK (Angeli & Valanides, 2009).

As hypothesized, TCoK positively predicts TPK and TPACK. In a training setting teachers may be required to collaborate. This could be achieved by establishing online learning communities (Zhang et al., 2019). In this way, teachers could directly experience the advantages of collaboration for their professional development and form positive attitudes towards it. Since attitudes are important in predicting Use, they may be addressed in professional development programs as well. To this end, the usefulness and ease of use for the individual (prospective) teacher may be demonstrated (Scherer et al., 2019).

7. Outlook

Scherer, Tondeur, Siddiq, and Baran (2018) showed that positive attitudes towards technology are positively associated with TPACK. Hence, besides a direct influence on Use, positive attitudes may also have an indirect positive influence via TPACK, TPK, or TCoK. To consider such indirect paths could be an avenue for future research. Moreover, we investigated technology integration across subjects, which may be a reasonable starting point (Sailer et al., 2021; this issue). Qualitative research would be helpful to identify how technology integration manifests itself in various subjects. Based on this, specific questionnaires for specific subjects, e.g., business education or language learning, could be developed.

Due to the quantitative empirical character of our study, we addressed the current technological state as delineated by the DigComp 2.1 framework. However, the potential of artificial intelligence in education has not been fully exploited (Luckin, Holmes, Griffiths, & Forcier, 2016). An example for educational technology driven by artificial intelligence are chatbots (Smutny & Schreiberova, 2020) or social robots that assist teachers (Papadopoulos et al., 2020; Guggemos, Seufert, & Sonderegger, 2020). The fundamental difference of such smart machines (devices driven by artificial intelligence) from conventional technology may be that smart machines are no longer be regarded as a tool, but as a partner (Daugherty & Wilson, 2018; Davenport & Kirby, 2016). Currently, TK addresses how to use technology. However, for smart machines, collaboration could be a more appropriate conception. Just as teachers need knowledge how to efficiently collaborate with human colleagues, they may need knowledge about collaborating with smart machines (concept of augmentation: Seufert, Guggemos, & Sailer, 2021). Daugherty and Wilson (2018) refer to such knowledge as *fusion skills*.

Assuming that fusion skills will become relevant in many areas and following our conceptual framework, teachers could be expected to foster those skills among their students, i.e., knowledge how to collaborate with smart machines. This may be well in line with initiatives in many countries to integrate computational thinking (Shute, Sun, & Asbell-Clarke, 2017; Wing, 2006) into curricula (Bocconi, Chioccariello, Dettori, Ferrari, & Engelhardt, 2016). Among others, computational thinking aims at enabling students to take benefit from artificial intelligence (Gadanidis, 2017; Repenning, Webb, & Ioannidou, 2010).

Credit author statement

Josef Guggemos: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. Sabine Seufert: Conceptualization, Investigation, Writing - review & editing, Resources, Supervision, Funding acquisition

Acknowledgment

This work was supported by the State Secretariat for Education, Research and Innovation, Switzerland, (project DigiCompToTeach). Our thanks go to the nine schools that participated in this study. Moreover, we would like to thank the reviewers for the elaborate and thoughtful comments on previous versions of this paper. All statements expressed in this article are those of the authors and do not reflect the official opinions or policies of the authors' host affiliations or the supporting institution.

Appendix

S1

Items to capture the constructs.

Construct	No.	Item	λ	M	SD
Attitudes Technology	1	I like working with digital media in class	.92	5.37	1.34
	2	The use of digital media is very useful for me as a teacher	.83	5.55	1.37
Attitudes Content	3	I like to address basic knowledge about digitization in class ('computer science knowledge': principles of digitization, universal technologies)	.92	4.84	1.72
	4	I like to expand my professional knowledge in terms of digital transformation	.94	5.13	1.61
TCoK	5	As part of a department, I am able to create a common vision and strategy for the digital development of the school	.85	4.12	1.74
	6	I am able to create new digital modules, materials, or examinations in collaboration with other teachers	.89	4.57	1.73
	7	I can use databases to exchange knowledge (collection of good practices)	.79	4.09	1.76
	8	I can help to shape the organizational framework of the school (e.g., digital curriculum development in a team)	.75	3.62	1.89
TK	9	I can use advanced search strategies (e.g., search operators) to perform a search query on the Internet	.71	5.18	1.58
	10	I can create and edit digital content in different formats (e.g., PDF and PPTX)	.69	5.66	1.51
	11	I can use digital communication tools appropriately (e.g., online chat, instant messaging, blogs)	.73	4.20	1.81
	12	I know what to do if my computer is infected by malware	.59	5.01	1.75
	13	I have various strategies and ways to efficiently find a solution for technical problems with digital media	.72	5.01	1.64
TPK	14	I can validly assess the competence of my students with digital tools (competence diagnostics)	.80	3.89	1.53
	15	I can adapt my teaching with digital media depending on the students' learning progress (individualization according to learning progress)	.81	4.19	1.59
	16	I know how to avoid disruptions when teaching with digital media (classroom management)	.85	4.21	1.51
	17	I can use a wide range of different digital teaching and learning methods	.81	4.09	1.67
	18	I can select teaching-learning forms with digital media in order to effectively promote students' thinking and learning processes	.87	4.14	1.57
TPACK	19	I can use digital media in the classroom to promote the ability of students to find, understand, and evaluate online information	.79	4.68	1.35
	20	I can promote the use of digital media in the classroom to enable students to edit and create digital content	.88	4.27	1.52
	21	I can promote that students work together appropriately by adequate means in digital media lessons	.86	4.23	1.47
	22	I can promote that students independently identify and solve problems using digital media	.80	4.03	1.56
	23	I can promote students' ability to use ICT applications (e.g., Microsoft Office ® package) in digital media lessons	.72	4.56	1.70
Use Technology	24	Individualization of lessons according to learning progress using digital media	.83	2.23	1.27
	25	Blended learning scenarios (e.g., Flipped Classroom)	.57	1.75	0.98
	26	Digital teaching-learning forms in lessons	.85	2.94	1.32
Use Content	27	Topics of digitization in my teaching	.79	2.65	1.14
	28	Cross-disciplinary teaching on digitization topics	.74	1.83	0.93
	29	Promote interdisciplinary competencies of learners in dealing with digital media (e.g., dealing with online information)	.77	2.58	1.13

Note. Knowledge and attitudes related items (1–23) measured on seven-point scales of rating ranging from "does not apply at all" to "applies very strongly". Use related items (24–29) measured with five-point scales of rating: never, infrequently (1–2 times per semester), occasionally (3–5 times per semester), frequently (every month), and very frequently (every week).

References

AERA, A. P. A., & NCME. (2014). *Standards for educational and psychological testing.* Washington D.C.: American Educational Research Association.

Agyei, D. D., & Voogt, J. M. (2011). Exploring the potential of the will, skill, tool model in Ghana: Predicting prospective and practicing teachers' use of technology. *Computers & Education, 56*(1), 91–100. https://doi.org/10.1016/j.compedu.2010.08.017

Agyei, D. D., & Voogt, J. (2012). Developing technological pedagogical content knowledge in pre-service mathematics teachers through collaborative design. *Australasian Journal of Educational Technology, 28*(4). https://doi.org/10.14742/ajet.827

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education, 52*(1), 154–168. https://doi.org/10.1016/j.compedu.2008.07.006

Backfisch, I., Lachner, A., Hische, C., Loose, F., & Scheiter, K. (2020). Professional knowledge or motivation? Investigating the role of teachers' expertise on the quality of technology-enhanced lesson plans. *Learning and Instruction, 66*, 101300. https://doi.org/10.1016/j.learninstruc.2019.101300

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Bell, A., Jones, K., & Fairbrother, M. (2018). Understanding and misunderstanding group mean centering: A commentary on kelley et al.'s dangerous practice. *Quality and Quantity, 52*(5), 2031–2036. https://doi.org/10.1007/s11135-017-0593-5

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A., & Engelhardt, K. (2016). *Developing computational thinking in compulsory education*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2791/792158

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Carretero, S., Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. Luxembourg: European Union (EU). https://doi.org/10.2760/38842

Christensen, R., & Knezek, G. (2008). Self-report measures and findings for information technology attitudes and competencies. In J. M. Voogt, & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 349–365). Berlin: Springer.

Claro, M., Salinas, A., Cabello-Hutt, T., San Martín, E., Preiss, D. D., Valenzuela, S., et al. (2018). Teaching in a Digital Environment (TIDE): Defining and measuring teachers' capacity to develop students' digital information and communication skills. *Computers & Education, 121*, 162–174. https://doi.org/10.1016/j.compedu.2018.03.001

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology, 25*(3), 325–334. https://doi.org/10.1007/s10869-010-9181-6

Darling-Hammond, L., Newton, X., & Wei, R. C. (2010). Evaluating teacher education outcomes: A study of the stanford teacher education programme. *Journal of Education for Teaching, 36*(4), 369–388. https://doi.org/10.1080/02607476.2010.513844

Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*. La Vergne: Harvard Business Review Press.

Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines*. New York, NY: Harper Business.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319. https://doi.org/10.2307/249008

Depaepe, F., & König, J. (2018). General pedagogical knowledge, self-efficacy and instructional practice: Disentangling their relationship in pre-service teacher education. *Teaching and Teacher Education, 69*, 177–190. https://doi.org/10.1016/j.tate.2017.10.003

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education, 69*, 485–492. https://doi.org/10.1016/j.compedu.2013.04.013

European Commission. (2019). *2nd survey of schools: ICT in education: Final report*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2759/23401

Farjon, D., Smits, A., & Voogt, J. M. (2019). Technology integration of pre-service teachers explained by attitudes and beliefs, competency, access, and experience. *Computers & Education, 130*, 81–93. https://doi.org/10.1016/j.compedu.2018.11.010

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA). https://www.iea.nl/publications/study-reports/preparing-life-digital-world.

Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: A comparison of four procedures. *Internet Research, 29*(3), 430–447. https://doi.org/10.1108/IntR-12-2017-0515

Gadanidis, G. (2017). Artificial intelligence, computational thinking, and mathematics education. *International Journal of Information and Learning Technology, 34*(2), 133–139. https://doi.org/10.1108/IJILT-09-2016-0048

Gil-Flores, J., Rodríguez-Santero, J., & Torres-Gordillo, J.-J. (2017). Factors that explain the use of ICT in secondary-education classrooms: The role of teacher characteristics and school infrastructure. *Computers in Human Behavior, 68*, 441–449. https://doi.org/10.1016/j.chb.2016.11.057

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education, 57*(3), 1953–1960. https://doi.org/10.1016/j.compedu.2011.04.010

Guggemos, J., Seufert, S., & Sonderegger, S. (2020). Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology, 51*(5), 1864–1883. https://doi.org/10.1111/bjet.13006

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*(3), 405–431. https://doi.org/10.1108/IMR-09-2014-0304

Hsu, L. (2016). Examining EFL teachers' technological pedagogical content knowledge and the adoption of mobile-assisted language learning: A partial least square approach. *Computer Assisted Language Learning, 29*(8), 1287–1297. https://doi.org/10.1080/09588221.2016.1278024

Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods, 51*(5), 2337–2355. https://doi.org/10.3758/s13428-018-1187-4

Kampylis, P., Punie, Y., & Devine, J. (2015). *Promoting effective digital-age learning: A European framework for digitally-competent educational organisations*. Luxembourg: European Union (EU). https://doi.org/10.2791/54070

Kirschner, P. A. (2015). Do we need teachers as designers of technology enhanced learning? *Instructional Science, 43*(2), 309–322. https://doi.org/10.1007/s11251-015-9346-9

Kirschner, P. A., & Bruyckere, P. de (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education, 67*, 135–142. https://doi.org/10.1016/j.tate.2017.06.001

Knezek, G., & Christensen, R. (2016). Extending the will, skill, tool model of technology integration: Adding pedagogy as a new model construct. *Journal of Computing in Higher Education, 28*(3), 307–325. https://doi.org/10.1007/s12528-016-9120-2

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.

Koehler, M. J., Mishra, P., & Cain, W. (2013). What is technological pedagogical content knowledge (TPACK)? *Journal of Education, 193*(3), 13–19. https://doi.org/10.1177/002205741319300303

Koh, J. H. L., Chai, C. S., & Tsai, C.-C. (2013). Examining practicing teachers' perceptions of technological pedagogical content knowledge (TPACK) pathways: A structural equation modeling approach. *Instructional Science, 41*(4), 793–809. https://doi.org/10.1007/s11251-012-9249-y

Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research, 77*(4), 575–614. https://doi.org/10.3102/0034654307309921

Loncke, J., Eichelsheim, V. I., Branje, S. J. T., Buysse, A., Meeus, W. H. J., & Loeys, T. (2018). Factor score regression with social relations model components: A case study exploring antecedents and consequences of perceived support in families. *Frontiers in Psychology, 9*, 1699. https://doi.org/10.3389/fpsyg.2018.01699

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. London: Pearson Education.

Mishra, P., & Koehler, M. J. (2006). Technological edagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Nistor, N. (2014). When technology acceptance models won't work: Non-significant intention-behavior effects. *Computers in Human Behavior, 34*, 299–300. https://doi.org/10.1016/j.chb.2014.02.052

Oberländer, M., Beinicke, A., & Bipp, T. (2020). Digital competencies: A review of the literature and applications in the workplace. *Computers & Education, 146*, 103752. https://doi.org/10.1016/j.compedu.2019.103752

OECD. (2015). *Students, computers and learning*. Advance online publication. https://doi.org/10.1787/9789264239555-en

Papadopoulos, I., Lazzarino, R., Miah, S., Weaver, T., Thomas, B., & Koulouglioti, C. (2020). A systematic review of socially assistive robots in pre-tertiary education. *Computers & Education*, 103924. https://doi.org/10.1016/j.compedu.2020.103924

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59–81. https://doi.org/10.1080/00461520.2016.1145550

Petko, D. (2012). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the 'will, skill, tool' model and integrating teachers' constructivist orientations. *Computers & Education, 58*(4), 1351–1359. https://doi.org/10.1016/j.compedu.2011.12.013

Petko, D. (2020). Quo vadis TPACK? Scouting the road ahead. *EdMedia + Innovate Learning, 23*, 1277–1286.

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. https://doi.org/10.3758/brm.40.3.879

Prenger, R., Poortman, C. L., & Handelzalts, A. (2019). The effects of networked professional learning communities. *Journal of Teacher Education, 70*(5), 441–452. https://doi.org/10.1177/0022487117753574

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/159770

Repenning, A., Webb, D., & Ioannidou, A. (2010). Scalable game design and the development of a checklist for getting computational thinking into public schools. In G. Lewandowski, S. Wolfman, T. J. Cortina, & E. L. Walker (Eds.), *Proceedings of the 41st ACM technical symposium on Computer science education - sigcse* (Vol. 10, p. 265). New York, New York, USA: ACM Press. https://doi.org/10.1145/1734263.1734357.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Ryan, R. M., & Weinstein, N. (2009). A self-determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*(2), 224–233. https://doi.org/10.1177/1477878509104327

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 561–582. https://doi.org/10.1080/10705510903203433

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with Fimix-Pls: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review, 63*, 34–62. https://doi.org/10.1007/BF03396886

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998–4010. https://doi.org/10.1016/j.jbusres.2016.06.007

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (tam): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education, 128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Scherer, R., Siddiq, F., & Tondeur, J. (2020). All the same or different? Revisiting measures of teachers' technology acceptance. *Computers & Education, 143*, 103656. https://doi.org/10.1016/j.compedu.2019.103656

Scherer, R., Tondeur, J., & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education, 112*, 1–17. https://doi.org/10.1016/j.compedu.2017.04.012

Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior, 80*, 67–80. https://doi.org/10.1016/j.chb.2017.11.003

Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education, 157*, 103967. https://doi.org/10.1016/j.compedu.2020.103967

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK). *Journal of Research on Technology in Education, 42*(2), 123–149. https://doi.org/10.1080/15391523.2009.10782544

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior, 115*. https://doi.org/10.1016/j.chb.2020.106552. In this issue.

Seufert, S., Guggemos, J., Tarantini, E., & Schumann, S. (2019). Developing and validating a framework of digital competencies for teachers in the commercial domain. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 115*(2), 312–339. https://doi.org/10.25162/ZBW-2019-0013

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Siddiq, F., Scherer, R., & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): A new construct in 21st century education. *Computers & Education, 92–93*, 1–14. https://doi.org/10.1016/j.compedu.2015.10.006

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education, 151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. *Review of Educational Research, 79*(2), 625–649. https://doi.org/10.3102/0034654308325896

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record, 104*(8), 1591–1626. https://doi.org/10.1111/1467-9620.00214

Teo, T., & van Schaik, P. (2012). Understanding the intention to use technology by preservice teachers: An empirical test of competing theoretical models. *International Journal of Human-Computer Interaction, 28*(3), 178–188. https://doi.org/10.1080/10447318.2011.581892

Tiede, J., Grafe, S., & Hobbs, R. (2015). Pedagogical media competencies of preservice teachers in Germany and the United States: A comparative analysis of theory and practice. *Peabody Journal of Education, 90*(4), 533–545. https://doi.org/10.1080/0161956X.2015.1068083

Tondeur, J., van Braak, J., Siddiq, F., & Scherer, R. (2016). Time for a new approach to prepare future teachers for educational technology use: Its meaning and measurement. *Computers & Education, 94*, 134–150. https://doi.org/10.1016/j.compedu.2015.11.009

Tseng, F.-C., & Kuo, F.-Y. (2014). A study of social participation and knowledge sharing in the teachers' online professional community of practice. *Computers & Education, 72*, 37–47. https://doi.org/10.1016/j.compedu.2013.10.005

Valtonen, T., Sointu, E., Kukkonen, J., Kontkanen, S., Lambert, M. C., & Mäkitalo-Siegl, K. (2017). TPACK updated to measure pre-service teachers' twenty-first century skills. *Australasian Journal of Educational Technology, 33*(3). https://doi.org/10.14742/ajet.3518

Van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & Haan, J. de (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577–588. https://doi.org/10.1016/j.chb.2017.03.010

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review, 15*, 17–40. https://doi.org/10.1016/j.edurev.2015.04.002

Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education, 61*, 47–59. https://doi.org/10.1016/j.tate.2016.10.001

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Venkatesh, V., Thong, J., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Voogt, J. M., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge - a review of the literature. *Journal of Computer Assisted Learning, 29*(2), 109–121. https://doi.org/10.1111/j.1365-2729.2012.00487.x

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology, 103*(4), 952–969. https://doi.org/10.1037/a0025125

Warwas, J., & Helm, C. (2018). Professional learning communities among vocational school teachers: Profiles and relations with instructional quality. *Teaching and Teacher Education, 73*, 43–55. https://doi.org/10.1016/j.tate.2018.03.012

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Zhang, S., Liu, Q., & Cai, Z. (2019). Exploring primary school teachers' technological pedagogical content knowledge (TPACK) in online collaborative discourse: An epistemic network analysis. *British Journal of Educational Technology, 50*(6), 3437–3455. https://doi.org/10.1111/bjet.12751

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206. https://doi.org/10.1086/651257