

Please quote as: Wambsganss, T., Janson, A., Söllner, M., Koedinger, K. & Leimeister, J. M. (2025). Improving Students' Argumentation Skills Using Dynamic Machine-Learning–Based Modeling. *Information Systems Research (ISR)*, 36, 474–507. doi: 10.1287/isre.2021.0615



## Information Systems Research

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Improving Students' Argumentation Skills Using Dynamic Machine-Learning—Based Modeling

Thiemo Wambsganss; , Andreas Janson; , Matthias Söllner; , Ken Koedinger; , Jan Marco Leimeister

To cite this article:

Thiemo Wambsganss; , Andreas Janson; , Matthias Söllner; , Ken Koedinger; , Jan Marco Leimeister (2025) Improving Students' Argumentation Skills Using Dynamic Machine-Learning—Based Modeling. Information Systems Research 36(1):474–507. <https://doi.org/10.1287/isre.2021.0615>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2024 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Improving Students' Argumentation Skills Using Dynamic Machine-Learning–Based Modeling

Thiemo Wambsganss,<sup>a,\*</sup> Andreas Janson,<sup>b</sup> Matthias Söllner,<sup>c</sup> Ken Koedinger,<sup>d</sup> Jan Marco Leimeister<sup>b,e</sup>

<sup>a</sup>Institute Digital Technology Management, Bern University of Applied Sciences, 3005 Bern, Switzerland; <sup>b</sup>Institute of Information Systems and Digital Business, University of St. Gallen, 9000 St. Gallen, Switzerland; <sup>c</sup>Research Center for IS Design, Information Systems and Systems Engineering, University of Kassel, 34121 Kassel, Germany; <sup>d</sup>Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213; <sup>e</sup>Research Center for IS Design, Information Systems, University of Kassel, 34121 Kassel, Germany

\*Corresponding author

Contact: thiemo.wambsganss@bfh.ch,  <https://orcid.org/0000-0002-7440-9357> (TW); andreas.janson@unisg.ch,  <https://orcid.org/0000-0003-3149-0340> (AJ); soellner@uni-kassel.de,  <https://orcid.org/0000-0002-1347-8252> (MS); koedinger@cmu.edu,  <https://orcid.org/0000-0002-5850-4768> (KK); janmarco.leimeister@unisg.ch,  <https://orcid.org/0000-0002-1990-2894> (JML)

Received: December 5, 2021

Revised: June 23, 2022; January 23, 2023;  
June 10, 2023; December 6, 2023;  
March 6, 2024; March 28, 2024

Accepted: April 11, 2024

Published Online in Articles in Advance:  
June 10, 2024

<https://doi.org/10.1287/isre.2021.0615>

Copyright: © 2024 The Author(s)

**Abstract.** Argumentation is an omnipresent rudiment of daily communication and thinking. The ability to form convincing arguments is not only fundamental to persuading an audience of novel ideas but also plays a major role in strategic decision making, negotiation, and constructive, civil discourse. However, humans often struggle to develop argumentation skills, owing to a lack of individual and instant feedback in their learning process, because providing feedback on the individual argumentation skills of learners is time-consuming and not scalable if conducted manually by educators. Grounding our research in social cognitive theory, we investigate whether dynamic technology-mediated argumentation modeling improves students' argumentation skills in the short and long term. To do so, we built a dynamic machine-learning (ML)–based modeling system. The system provides learners with dynamic writing feedback opportunities based on logical argumentation errors irrespective of instructor, time, and location. We conducted three empirical studies to test whether dynamic modeling improves persuasive writing performance more so than the benchmarks of scripted argumentation modeling (H1) and adaptive support (H2). Moreover, we assess whether, compared with adaptive support, dynamic argumentation modeling leads to better persuasive writing performance on both complex and simple tasks (H3). Finally, we investigate whether dynamic modeling on repeated argumentation tasks (over three months) leads to better learning in comparison with static modeling and no modeling (H4). Our results show that dynamic behavioral modeling significantly improves learners' objective argumentation skills across domains, outperforming established methods like scripted modeling, adaptive support, and static modeling. The results further indicate that, compared with adaptive support, the effect of the dynamic modeling approach holds across complex (large effect) and simple tasks (medium effect) and supports learners with lower and higher expertise alike. This work provides important empirical findings related to the effects of dynamic modeling and social cognitive theory that inform the design of writing and skill support systems for education. This paper demonstrates that social cognitive theory and dynamic modeling based on ML generalize outside of math and science domains to argumentative writing.

**History:** Gautam Pant, Senior Editor; Hong Guo, Associate Editor.



**Open Access Statement:** This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Information Systems Research. Copyright © 2024 The Author(s). <https://doi.org/10.1287/isre.2021.0615>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

**Funding:** T. Wambsganss acknowledges funding by a SNSF Doc.Mobility grant [Grant 200207] to collaborate with K. Koedinger from Carnegie Mellon University. J. M. Leimeister [Grant 192718] and A. Janson [Grant 221281] also acknowledge funding from the SNSF. The authors furthermore thank the German Federal Ministry of Education and Research [BMBF, Grant 16DHBK1073] for partly funding this research.

**Supplemental Material:** The online appendix is available at <https://doi.org/10.1287/isre.2021.0615>.

**Keywords:** dynamic argumentation feedback • artificial intelligence for education • adaptive argumentation learning • adaptive learning • argumentation skills

## Introduction

Information is readily available in today's world, and the ability to memorize existing information is no longer prioritized. Rather, the skills and abilities required to structure and process knowledge are gaining importance. Job profiles have thus shifted toward interdisciplinary, ambiguous, and creative tasks (vom Brocke et al. 2018). Educational institutions must evolve their curricula when it comes to the composition of skills and knowledge conveyed (Jung and Lehrer 2017, Topi 2018). Skills such as critical thinking, collaboration, or problem solving have become particularly important (Fadel et al. 2015). This has already been recognized by institutions such as the Organization for Economic Cooperation and Development (OECD), which included this class of skills as a major element of its Learning Framework 2030 (OECD 2018). One example of these competencies is the ability to advance structured, reflective, and well-formed arguments (Toulmin 2003, Walton et al. 2008, Visser et al. 2022). Argumentation not only is essential to daily communication and thinking but also contributes significantly to multiple skills, such as communication, collaboration, and problem solving (Kuhn 1992). In studies dating back to Aristoteles, the ability to form convincing arguments has been recognized as essential to persuading audiences of novel ideas' merits and as playing major roles in analyzing different standpoints and in constructive or logical civil discourse for example, assessing whether news items are fake (Diana 2018, Deng et al. 2019).

However, teaching argumentation skills in professional organizations or educational institutions is hindered by a lack of individual and personal learning experiences (see, e.g., Vygotsky 1980, Hattie and Timperley 2007, Gupta and Bostrom 2013, Santhanam et al. 2016). Learners are frequently confronted with unfavorable educator-student ratios, not only in the distance-learning scenarios of massive open online courses MOOCs (Seaman et al. 2018) but also in traditional large-scale lectures on university campuses. Large classes in high schools, mass lectures at universities with more than 100 students per lecturer, and MOOCs with more than 1,000 participants impede individual interaction between learners and educators (Seaman et al. 2018, Winkler et al. 2021). The growth of MOOCs in recent years has fostered this development. In 2017, 33.1% of learners worldwide took at least one course online compared with 24.8% in 2012 (Lederman 2018). With the leading MOOC provider Coursera's launch of a successful initial public offering in March 2021, further expansion of distance learning settings is predicted (Tse and Roof 2021). Several studies have revealed that this lack of individualized support leads to procrastination, low learning outcomes, high dropout rates, and dissatisfaction with the overall learning experience

(Eom et al. 2006, Lehmann et al. 2014, Brinton et al. 2015, Hone and El Said 2016, Huang et al. 2021).

Information systems (IS) have been leveraged as technology-mediated learning tools over several decades to meet scalable and individual learning challenges (Leidner and Jarvenpaa 1995, Alavi and Leidner 2001, Gupta and Bostrom 2009). IS have been designed and developed to support individuals in learning how to argue in online debates (Wang et al. 2020), in collaborative learning settings (Dillenbourg et al. 2009, Gupta and Bostrom 2013), or through the provision of tailored, also called adaptive, argumentation feedback (Kulik and Fletcher 2016, Wang et al. 2020, Afrin et al. 2021). In fact, for at least 35 years (Smolensky et al. 1988), researchers have developed technology-enhanced argumentation support systems in various domains, including law (Pinkwart et al. 2009, Weber et al. 2024), science (Suthers and Hundhausen 2001, Osborne et al. 2016), and conversational argumentation (De Groot et al. 2007, Wambsganss et al. 2021). Nevertheless, the design and adoption of current argumentation learning systems lag behind recent technological developments in IS and natural language processing (NLP) research (Stab and Gurevych 2017b, Lawrence and Reed 2019, Afrin et al. 2021). This has resulted in a lack of interdisciplinary literature that investigates the effect of automated dynamic modeling on students' argumentation skills to enable technology-mediated learning (Huang et al. 2021, Xu et al. 2021). Therefore, it is not only learners in IS education who still struggle to develop argumentation skills owing to a lack of intelligent and instant evaluation in their individual learning process because current approaches fail to provide scalable and formative modeled feedback on learners' argumentation skills (Scheuer 2015, Lawrence and Reed 2019). In fact, a socio-technical IS perspective to support the formation of a holistic learner-centered argumentation learning system based on behavioral modeling feedback with the opportunity for learners to monitor and evaluate themselves based on advances in NLP and machine learning (ML) is missing (Scheuer et al. 2010, Huang et al. 2016, Lawrence and Reed 2019, Head et al. 2021, Xu et al. 2021). Past research has focused mostly on the provision of argumentation scripts (a.k.a. scripted modeling) or static modeling through representational guidance, which have become the standards to provide technology-mediated learning support at scale for argumentation (Scheuer et al. 2012, Fischer et al. 2013, Noroozi et al. 2020).

Although existing studies have extended the understanding of technology-mediated modeling feedback for argumentation learning, they have usually sought to address a general system design rather than specific research questions to investigate the underlying effects (see, e.g., Lippi and Torroni 2016, Chernodub et al. 2019, Wang et al. 2020, Afrin et al. 2021). Although earlier

studies made contributions in exploring the development and use of dynamic argumentation modeling systems (see, e.g., Chernodub et al. 2019, Lauscher et al. 2019, Wang et al. 2020, Wambsganss et al. 2022b), they often neglected to present experimental studies in real learning environments—particularly for the class of ML-based support systems—for further developments in the field. Still, scripting students with upfront argumentation structures has been the proven and evaluated approach to support students in learning how to argue in large-scale scenarios in the field (Dillenbourg and Hong 2008, Weinberger et al. 2010, Stegmann et al. 2012, Fischer et al. 2013). The same holds true for the broader IS literature concerned with technology-mediated learning that draws mainly upon approaches that are rather static when providing models of learner behavior (Gupta and Bostrom 2013, Sullivan et al. 2022). With the rise of NLP and ML, the class of educational systems that provides dynamic models of learner behavior and related argumentation feedback based on recent advances in argumentation mining has received special attention (Chernodub et al. 2019, Deng et al. 2019, Lauscher et al. 2019, Wang et al. 2020, Afrin et al. 2021). Argumentation mining is a subdiscipline of NLP that deals with the identification and classification of argumentative discourse structures in natural language (see, e.g., Lawrence and Reed 2019). It enables modeling of desired learner behavior so that learners can learn from a model in the sense of observational learning processes. However, besides the obvious benefits of argumentation mining for dynamic behavioral modeling, these systems have been rather poorly evaluated in controlled environments in comparison with the benchmarks of 1) scripted argumentation modeling and 2) adaptive argumentation support as well as not 3) not being experimentally examined in a long-term field experiment to measure the effect of dynamic modeling in a real-world learning scenario on students' argumentation skills.

Drawing on the core principles of social cognitive theory (Bandura 1986, 2001), we propose the design of a theory-driven solution of a dynamic argumentation modeling system based on ML that models learner behavior in a dynamic way to provide individuals with transparent, individual, and adaptive feedback based on their logical argumentation errors (Rosé et al. 2008). With a dynamic modeling system, we imply a learning tool that provides a model of desired learner behavior that provides individual and personalized feedback on the behavioral aspects of learning (e.g., argumentation skills). In our instantiation, we aim not only to address the educational task (writing coherent arguments) but also to build a holistic argumentation learning tool embedded in a pedagogical scenario (structure), focusing on the learner through a user-centered and theory-motivated design (Gupta and Bostrom 2013, 2009).

Based on behavioral modeling as a cornerstone of social cognitive theory (Bandura 2001), we suggest that dynamic behavioral modeling of argumentation and corresponding feedback improves the ability of learners to improve their argumentation and, thus, leads to higher argumentation skill learning outcomes. To evaluate the impact of dynamic argumentation modeling feedback on humans' argumentation skills, we conducted three empirical studies. We demonstrated and evaluated the proposed theory-driven learning system by testing 1) whether *dynamic argumentation modeling improves persuasive writing performance more so than scripted argumentation modeling* (H1—study 1) and 2) *more so than adaptive support approaches* (H2—study 2) in controlled laboratory experiments. Moreover, we test 3) *whether, compared with adaptive support, dynamic argumentation modeling leads to better persuasive writing performance on both complex and simple tasks* (H3). (H3—study 2). Finally, we investigate whether 4) *dynamic modeling on repeated argumentation tasks (over three months) leads to better learning in comparison with static modeling* (H4—study 3). We measure persuasive writing performance through two variables: *objective quality of argumentation*, assessed in the formal coherences of arguments according to Toulmin (1984), and *subjective quality of argumentation*, assessed through external ratings. Argumentation skill learning is measured through learners' persuasive writing performance in a different domain after a longer period of time (three months). Our results indicate that dynamic modeling leads to a stronger increase in learners' argumentation skills (when considering objective argumentation), even with a transfer effect to other argumentative domains. The findings show that dynamic behavioral modeling significantly improves learners' argumentation skills across all conditions, outperforming established methods like the benchmarks of scripted modeling (study 1), traditional adaptive support (study 2), and static modeling (study 3). Our dynamic modeling approach seems to be able to adapt to task difficulty (study 2) and learner expertise (across all studies) robustly. It supports students with low and high expertise as well as when completing easy and hard tasks. Regarding task difficulty, we additionally observe an even stronger supporting effect of our approach when students are faced with complex tasks (compared with simple tasks). Furthermore, our third study shows that dynamic modeling based on ML helps students to better train their argumentation skills compared with static modeling and no modeling over a longer period of time.

Our research has several contributions. We demonstrate and evaluate the effectiveness of dynamic ML-based modeling on students' short-term argumentation skills by rigorously comparing our system with the current benchmarks of scripted modeling (study 1), traditional adaptive support (study 2), and by comparing

dynamic and static modeling in a real-world learning setting over the course of three months (study 3). This is especially novel, whereas past work has developed mainly theories of learning by examples and by doing (e.g., ACT-R theory and SimStudent theory) in STEM context (science, technology, engineering, and mathematics subjects). This paper demonstrates that this theory generalizes outside of math and science domains to argumentative writing (see, e.g., Anderson 1986, Anderson et al. 1995, Matsuda et al. 2015). The results demonstrate how NLP may be leveraged in designing ML-based dynamic learning systems that provide ongoing formative learning support throughout the learning journey. Hence, we contribute to social cognitive theory by demonstrating how to provide adaptive and personalized models as a source for observational learning processes throughout a student's learning journey cycle. Finally, our results exemplify how skills may be supported in a scalable and individual way in large-scale scenarios. Thus, we establish the foundation for other researchers and educators to design similar tools aimed at supporting learner skills.

## Related Work

### Argumentation Skills and Underlying Theoretical Models

Argumentation is an omnipresent rudiment of daily communication and thinking (Kuhn 1992). In general, argumentation aims to increase or decrease a controversial standpoint's acceptability (van Eemeren et al. 1996). Logically, structured arguments are a required precondition for persuasive conversations, general decision making, and drawing acknowledged conclusions (Duschl and Osborne 2002). Across multiple fields, including law, science, politics, or management, individuals must support their claims with essential facts, argue in support of conclusions derived from those facts, and counter their opponent's claims in a principled way to convince others of their position or justify a conclusion (von Aufschnaiter et al. 2008). Over several decades, research has demonstrated that humans are generally deficient in argumentation (see, e.g., Tversky and Kahneman 1974, Marcus and Rips 1979, Byrne 1989); they often fail to recognize the difference between merely expressing an opinion and making a fact-based claim. Moreover, they do not rebut others' arguments but ignore points of conflict and persist with their own arguments (Byrne 1989). Thus, research in fields such as human-computer interaction (HCI) or IS has demonstrated increased interest in developing argumentation systems to support individuals (van Eemeren et al. 1996).

Argumentation theories have a long history in philosophy, linguistic research, and mathematics. Aristotle provided one of the first foundations in his theory of

persuasion. Many frameworks and "rules" of argumentation have since been proposed and identified (Kuhn 1992, Toulmin 2003, Walton et al. 2008). In his fundamental theory, Aristotle distinguished between three interconnected principles of persuasion—*logos* (logic and proof of argumentation), *ethos* (authority and credibility of the speaker), and *pathos* (empathy and vivid language)—on which an individual can build his or her effort to persuade an opponent. *Logos* focuses on the general formality and structure of argumentation, *ethos* depends largely on the individual and his or her relationship to the opponent, and *pathos* concerns context-related emotions and the strength of languages.

Formal logic, a branch of mathematics, has long been the main perspective on persuasive argumentation (see, e.g., North Whitehead and Russell 1910, Scheuer et al. 2010). More recently, however, theories have also focused on practical human argumentation (see, e.g., Toulmin 1984, Kuhn 1992). Most theoretical and practical approaches to argumentation vary in their level of detail, perspective, and specific context of applicability. Nevertheless, several scholars (see, e.g., Scheuer et al. 2010, 2012; Scheuer 2015) have observed that logic is likely the foundation that underlies different theoretical argumentation approaches. Argumentation theory generally agrees on the importance of *logos* as a basis for proper argumentation. Argumentation discourse should consider all relevant facts, claims should be well-grounded and supported by premises, and both supporting and conflicting claims should be taken into account (Toulmin 2003).

Most argumentation theories formulate formal argumentation models that address the logical component of Aristotle's theory of persuasion. The Toulmin model is among the most prominent of these (Toulmin 1984, 2003) and has been applied in IS research (e.g., for recommendation agents, see Kim and Benbasat 2006). Accordingly, an argument consists of several components, including a claim and at least one premise. The claim is the central component and statement, which is justified by premises. According to Toulmin's argumentation theory, a *good* argument involves a logical structure founded on grounds, claim, and warrant, whereas the grounds are the evidence used to prove a claim (Toulmin 1984). For example, according to the Toulmin Model, each argument can be broken down into six parts (claim, premise/ground, warrant, backing, rebuttal, and qualifier). Each component fulfills a specific argumentative role and complements the others to create the argument. Nevertheless, *claim* and *premise* (also called *evidence*) are generally considered the main components of every argument, and the rest are supporting subargument parts that may or may not exist in an argument (Figure 1; also see Scheuer 2015 and Stab and Gurevych 2017b):

**Figure 1.** Basic Argumentation Model with Claim and Premise

- **Claim:** the central point or conclusion of the argument. The goal is to convince the opponent of the claim's truthfulness.
- **Premise (fact, evidence, data):** the support or rationale for the claim; the data or evidence that the arguer uses to explain and support his or her claim.

### Technology-Mediated Argumentation Learning Systems

Based on argumentation theory and, most prominently, the Toulmin model, researchers have developed different ISs to support and teach argumentation; over a period of 35 years, more than 60 have been developed and published to support learners in creating, editing, interpreting, or reviewing arguments (Smolensky et al. 1988, Suthers and Hundhausen 2001, Pinkwart et al. 2009, Huang et al. 2016, Chernodub et al. 2019, Wang et al. 2020, Afrin et al. 2021).

The paradigms of computer-supported collaborative learning (CSCL; Koschmann 1996, Dillenbourg et al. 2009) and intelligent tutoring systems (ITS) in the field of educational technology are particularly relevant to argumentation learning (Scheuer 2015) because argumentative discussions and debates have been identified as a key factor for collaborative learning settings in particular. Therefore, argumentation has emerged as a focus area in CSCL. Research on ITS centers more on analyzing, modeling, and supporting technology-enhanced learning activities in specific domains (Dillenbourg 2002, Nye 2014). The combination of CSCL and ITS to support collaboration and argumentation both adaptively and individually is a relatively new research area (Fischer et al. 2013). Researchers have designed and evaluated several tools based on input masks and representational guidelines to support high school students' active writing processes. This has been investigated across various fields, including law (Pinkwart et al. 2009), science (Suthers and Hundhausen 2001, Osborne et al. 2016), and conversational argumentation (De Groot et al. 2007). Following Scheuer (2015), and with the lenses of social-cognitive theory (Bandura 1986, 2001), three different IT-based argumentation learning systems can be distinguished: *static modeling through representational guidance*, *scripted modeling*, and *adaptive support approaches* (presented in Table 1). Especially script-based approaches and static modeled feedback (e.g., knowledge of results feedback or example-based learning) have developed into the

practical standard for argumentation learning at scale because pre-scripting students' argumentation structures is domain independent, scalable, and effective for argumentation skill learning (Fischer et al. 2013, Scheuer 2015, Noroozi et al. 2020). Furthermore, adaptive support approaches have been built to provide students with adaptive feedback, for example, based on dashboard-like scores (Lauscher et al. 2019) or feedback messages (Scheuer et al. 2012). Nevertheless, these approaches are often not evaluated with learners and miss an entire pedagogical perspective (Scheuer 2015). We propose a system based on dynamic argumentation modeling centered around a socio-technical design perspective (Bostrom and Heinen 1977). We aim not only to address the educational task (writing coherent arguments) but also to build a holistic argumentation learning system embedded in a pedagogical scenario (structure), focusing on the learner through a user-centered and theory-motivated design.

Moreover, we built the class of argumentation learning systems on social cognitive theory (Bandura 1986, 2001) using recent advances in ML, which enable an adaptive and individual learning experience irrespective of instructor, time, and place (Xu et al. 2021). Thus, we consider our dynamic argumentation modeling based on ML as a subclass of adaptive support approaches (Scheuer et al. 2010, Scheuer 2015) that provide intelligent and personalized feedback compared with existing adaptive argumentation support systems.

Our approach builds on recent advances in NLP and ML, particularly argumentation mining. Argumentation mining is a research field in computational linguistics that focuses on the extraction and analysis of arguments from textual corpora as well as on following and analyzing the lines of argumentation (i.e., the interplay between arguments). Since 2007, scientists have published studies on argumentation mining in legal texts, online reviews, or debates (Palau and Moens 2009, Mochales and Moens 2011). Argumentation mining analyzes the arguments of a given text based on a defined argumentation structure (often based on Toulmin 2003). Argumentation structures can be identified on three different levels. First, a sentence containing an argument is identified to differentiate argumentative from nonargumentative text units (Florou et al. 2013). The second level involves classifying argument components into claims and premises (Mochales and Moens 2011, Stab and Gurevych 2014). The third level is the

**Table 1.** Overview of Different IT-Based Argumentation Learning Approaches from the Perspective of Social-Cognitive Theory

| Study                         | Description                                                                                                                                               | Findings                                                                                                                                                                                                                                                                                                                                                                                                                                           | Theoretical lens                                                                                  | General approach                                                                                                                                                                                  | Limitations                                                                                                                                                                               |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Suthers and Hundhausen (2001) | Provide learners with hypotheses-evidence relations represented within the Belvedere system as node-and-link graph and table.                             | Different representations of argumentation structures guide argumentation skill learning of students.                                                                                                                                                                                                                                                                                                                                              | Static Modeling Through Representational-Guidance Theory (see, e.g., Suthers and Hundhausen 2003) | Try to support argumentation learning by providing representations of argumentation structures with the objective of stimulating and improving individual reasoning, collaboration, and learning. | Focus on the static representation of argumentation, not on the dynamic learning process<br>Larger argumentation graphs tend to confuse users and are not useful for longer writing tasks |
| Nussbaum et al. (2007)        | Provide learners with a special-purpose diagramming format (argumentation vee diagrams) to support consideration and integration of different viewpoints. | Graphic argumentation organizer resulted in more refutations of counterarguments. However, static instructions resulted in better integration of argument and counterargument (with stronger rebuttals and more balanced reasoning). Results indicate that representational notations can have significant effects on learners' interactions and may differ in their influence on subsequent collaborative use of the knowledge being manipulated. |                                                                                                   | One common approach is to represent argumentation discourse in the form of linked graphs.                                                                                                         | Representations are usually built on domain-dependent ontologies and thus limited for a wide-scale adoption (Scheuer et al. 2010)                                                         |
| Suthers and Hundhausen (2003) | Provide learners with different representational notations and compare their effects.                                                                     | Results indicate that representational notations can have significant effects on learners' interactions and may differ in their influence on subsequent collaborative use of the knowledge being manipulated.                                                                                                                                                                                                                                      |                                                                                                   |                                                                                                                                                                                                   |                                                                                                                                                                                           |
| Fischer et al. (2013)         | Support learners in the creation of argumentative texts through a form-like interface.                                                                    | Upfront argumentation scripting improves objective quality of argumentation.                                                                                                                                                                                                                                                                                                                                                                       | Scripted modeling (Dillenbourg 2002, Dillenbourg et al. 2009, Fischer et al. 2013)                | Aim to provide structured elements for argumentation learning processes to foster interactions based on script theory of guidance.                                                                | Missing adaptivity in scripted learning (Scheuer 2015)<br>Not necessarily applicable for longer argumentative writing tasks                                                               |
| Ravenscroft et al. (2010)     | Support learners with a Web-based tool supporting collaborative argumentation and other forms of real-time learning dialogue.                             | Scripted learning dialogue stimulates reasoning, collaborative thinking, and the development of argumentative dialogue practices.                                                                                                                                                                                                                                                                                                                  |                                                                                                   | One common approach is to let learners choose from predefined sentence openers when composing new text content.                                                                                   | Might limit the creativity of argumentation (i.e., overscripting; see, e.g., Dillenbourg 2002, Dillenbourg et al. 2009)                                                                   |
| McAlister et al. (2004)       | Support learners with sentence opener to facilitate a discussion with the system AcademicTalk.                                                            | Sentence openers encourage critical engagement with the opinions of others as well as the use of evidence and reasons to support claim.                                                                                                                                                                                                                                                                                                            |                                                                                                   | The standard of supporting argumentation skills at scale around a wide range of domains.                                                                                                          |                                                                                                                                                                                           |

Table 1. (Continued)

| Study                  | Description                                                                                                           | Findings                                                                                                                                                                                                                                                                                                                                                                | Theoretical lens                                                                                               | General approach                                                                                                                                                                                                                                                                                                                               | Limitations                                                                                                                                                                                                                                                                              |
|------------------------|-----------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Scheuer et al. (2014)  | Provide users with feedback messages based on predefined detected patterns.                                           | Graphical argument representations with feedback make argument structures explicit, encourage reflection on basic concepts of argumentation, reduce cognitive load, help to systematically explore a debate space, facilitate the evaluation of arguments, and serve as resources and stimuli for discussions.                                                          | Adaptive support (see, e.g., Scheuer et al. 2012, Scheuer 2015)                                                | Aim to provide textual pedagogical adaptive feedback on a learner's action and solutions, hints, and recommendations to encourage and guide future activities in the writing processes or automated evaluation to indicate whether an argument is syntactically and semantically correct.                                                      | Limited in providing adaptive and individual learning feedback<br>Lack of interdisciplinary socio-technical design perspective (e.g., learners' perspective).<br>Non-ML-based adaptive systems exist, (see, e.g., (Suthers and Hundhausen 2001, Pinkwart et al. 2009, Huang et al. 2016) |
| Chemodub et al. (2019) | Providing general users with the tool ArgumentTarger based on different ML models for claim and premise highlighting. | No learner/user evaluation.                                                                                                                                                                                                                                                                                                                                             |                                                                                                                | Feedback in these systems is usually based on prewritten text messages or meters to support meta cognition (see, e.g., McLaren et al. 2010, Scheuer et al. 2014).                                                                                                                                                                              | ML-based systems motivated and instantiated but not evaluated (see, e.g., Stab and Gurevych 2017b, Chemodub et al. 2019, Lauscher et al. 2019)                                                                                                                                           |
| Lauscher et al. (2019) | Supporting general users with the tool ArguminSci for argument claim detection in scientific writing.                 | No learner/user evaluation.                                                                                                                                                                                                                                                                                                                                             |                                                                                                                |                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                          |
| This study             | Supporting learners through dynamic ML-based modeling, enabling adaptive feedback.                                    | Dynamic behavioral modeling significantly improves learners' argumentation skills across domains, outperforming established methods like scripted modeling, adaptive support, and static modeling. Empirical findings related to the effects of dynamic modeling and social cognitive theory that inform the design of writing and skill support systems for education. | Dynamic modeling approach based on Social Cognitive Theory (dynamic behavioral modeling based on Bandura 1986) | We aim to provide and test a socio-technical and user-centered information systems built on dynamic ML-based modeling using NLP and ML. Theory-driven system based on ML that models learner behavior in a dynamic way to provide individuals with transparent, individual, and adaptive feedback based on their logical argumentation errors. | Domain dependent (depends on the used corpus).<br>Large implementation effort necessary before first application.<br>Accuracy and potentially erroneous feedback.                                                                                                                        |

identification of argumentative relations (Palau and Moens 2009, Stab and Gurevych 2014). Argumentation mining researchers have become increasingly interested in intelligent writing assistance (Song et al. 2014, Stab and Gurevych 2017b) because it enables modeling of desired learner behavior so that learners are able to learn from a model in the sense of observational learning processes. However, the complexity of this technology's use in pedagogical scenarios for educational purposes has hitherto been poorly assessed (Lawrence and Reed 2019).

In fact, in the technical disciplines of NLP, other forms of writing support systems exist. Since 2016, research from NLP has utilized argumentation mining algorithms and embedded these approaches in simple text input tools. Lippi and Torroni (2016) developed the first online argumentation mining tool that was made available to a broad audience. Their tool is available as a Web application and processes text that is input in the corresponding editor field. After processing and analyzing the text, the results are displayed on the user interface. Claims are displayed in bold font, whereas premises are displayed in italic style (Lippi and Torroni 2016). Also, Lauscher et al. (2019) have provided a simple text-input interface called ArguminSci that aims to support the holistic analyses of scientific publications, including the identification of argumentative components. Moreover, Chernodub et al. (2019) designed a system to make argumentation identification in text available for individuals. In their system, called TARGER, a user can analyze the argumentative structure of an input text. The results are then presented below the input, whereas claims are highlighted in red, and premises are marked in green. These approaches originated in the NLP community and were built mostly with the aim of making novel argumentation mining algorithms available for nonprograms (Lippi and Torroni 2016, Chernodub et al. 2019, Lauscher et al. 2019, Lawrence and Reed 2019). Nevertheless, they were not designed for a certain pedagogical scenario, nor with a learner-centered perspective. Furthermore, these systems do not take a technology-mediated learning perspective into account and are thus not necessarily easy to use and easy to access for students in their learning journey, because a user would always have to open the website, select a certain model, and then copy his or her text into the input field (Afrin et al. 2021, Xu et al. 2021, Noetel et al. 2022). All of these systems have in common that they are usually not empirically—neither qualitatively nor quantitatively—evaluated from an interdisciplinary perspective based on insights from educational technology, HCI, and NLP (Stab and Gurevych 2017b, Chernodub et al. 2019, Lawrence and Reed 2019, Wang et al. 2020, Afrin et al. 2021). In this regard, IS research can offer a promising viewpoint for investigating and evaluating a certain IS from a technology-mediated learning

perspective (Gupta and Bostrom 2009), incorporating the different disciplinary perspectives in the design, demonstration, and evaluation of the IS (Sidorova et al. 2008).

Despite the availability of literature on argumentation systems, broader insights of the impact of these systems, particularly for the new class dynamic argumentation modeling systems, are scarce (Scheuer et al. 2010, Wang et al. 2020, Afrin et al. 2021). Although existing studies have extended the understanding of technology-mediated argumentation learning, they were usually conducted to answer specific research questions rather than to obtain information that could be used for widespread application (see, e.g., Lippi and Torroni 2016, Chernodub et al. 2019, Wang et al. 2020, Afrin et al. 2021). As Stab and Gurevych (2017b, p. 649) stated, “It is still unknown, however, if feedback provides an adequate guidance for improving students’ argumentation skills. To answer this question, it is required to integrate the proposed model in writing environments and investigate the effect of different feedback types on the argumentation skills of students in future research.” Scheuer (2015, p. 126) also noted, “Rigorous empirical research with respect to adaptation strategies is almost absent; a broad and solid theoretical underpinning, or theory of adaptation for collaborative and argumentative learning is still lacking.” We aim to fill this gap by grounding our research on a social cognitive view of technology-mediated learning (Bandura 1986, 2001; Gupta and Bostrom 2013, 2009). Hence, we derive four hypotheses and test these hypotheses based on the theory-driven design of a natively instantiated dynamic argumentation modeling system.

### Social Cognitive View on Technology-Mediated Learning

When considering the related work, especially in the IS discipline, social cognitive theory provides a long-standing history on the mechanisms for learning with technology. Social cognitive theory was developed by Bandura as a theory that relates to learning, that is, acquiring knowledge through the observation of models (Bandura 1991, Compeau and Higgins 1995, Yi and Davis 2003, Sullivan et al. 2022). Behavioral modeling is a cornerstone of this theory and has been studied before in information systems research and relates mainly to two learning processes (see Gupta and Bostrom 2013): observation of others’ actions (vicarious learning processes; Sullivan et al. 2022) as well as learning by doing and experiencing consequences (enactive learning; Gupta and Bostrom 2013). As Bandura (1986) stated, “Behavior is, therefore, a product of both self-generated and external sources of influence” (p. 454). With respect to external sources, social cognitive theory does not restrict these sources to sources such as fellow peers and teachers. Rather, the theory

includes all external sources that model learning. These sources could relate to abstract curricula and video recordings but also digital sources (see, e.g., Zimmerman and Schunk 2001, van Gog and Rummel 2010). The latter is especially shown through IS research drawing upon social cognitive theory as the kernel theory for grounding the effects of digital learning environments and learning interventions (see, e.g., Compeau and Higgins 1995; Santhanam et al. 2008, 2016; Gupta and Bostrom 2013; Leung et al. 2022). The same holds true for our argumentation learning environment that models as a social agent desired learner behavior and contributes to vicarious learning processes. A key underlying theoretical mechanism that behavioral modeling provides to learners is observational learning (Schunk 2012, Sullivan et al. 2022). Prior research has especially focused on the provision of static models for learning processes, such as providing video instructions or static examples on how to accomplish a task (see, e.g., Gupta and Bostrom 2013, Sullivan et al. 2022). Nonetheless, we argue that despite the positive results of existing findings, a dynamic approach to behavioral modeling and observational learning processes may be superior. Thus, we outline in the next section our key hypotheses that guide the subsequent investigation of the effects of a dynamic behavioral modeling approach compared with existing approaches.

## Hypotheses Development

In this section, we will outline how we develop the hypotheses to investigate and demonstrate our theory-driven solution of dynamic behavioral modeling of argumentation to support individuals with adaptive and individual support in learning how to argue. The hypotheses were derived based on literature on (argumentation) skill learning (see, e.g., Pinkwart et al. 2009, Jonassen and Kim 2010, Wang et al. 2020), argumentation theories (Kuhn 1993, Toulmin 2003, Walton et al. 2008), technology-mediated learning with a focus on social cognitive theory (Alavi and Leidner 2001, Hattie and Timperley 2007, Gupta and Bostrom 2009, Han et al. 2021), as well as literature on argumentation mining, and the design of educational learning tools (see, e.g., Mayer 2014 and Santhanam et al. 2016).

### Dynamic Argumentation Modeling vs. Scripted Argumentation Modeling

When thinking about the provision of supporting learners in argumentation at scale, typically, two major approaches could be utilized: scripted argumentation modeling (i.e., providing upfront scripting for argumentation) or modeling learner behavior in a dynamic way and providing opportunities for observational learning through feedback, that is, showing desired learner behavior related to the topic of how to form arguments. Because the two concepts rely on different

theoretical mechanisms, and prior research does not provide conclusive insight on their effectiveness for argumentation learning in the business domain (Scheuer 2015, Noroozi et al. 2020), we disentangle these concepts in the following hypotheses development.

Concerning the first option, scripted modeling approaches provide learners with external facilitation and structures in forming arguments for discussing and communicating with others. Fischer et al. (2013) proposed a theoretical foundation for discussion, scripting approaches by deriving their theory of guidance. Their underlying hypothesis assumes that learning is facilitated by mental knowledge structures, which they call internal collaboration scripts. Fischer et al. (2013) defined different components of these internal scripts (play, scenes, roles, and scriptlets), and a set of principles for each of these components, for example, the script configuration principle. Students are provided with text fields to enter a claim, grounds, and qualification based on the Toulmin model of argumentation (Toulmin 1984). Several empirical studies showed that this pre-structuring of argumentation improves the objective quality of argumentation (see, e.g., Stegmann et al. 2012, Fischer et al. 2013). Based on these principles, researchers investigated different learning tools to improve the quality of argumentation through structured communication interfaces. The interventions were mostly on the scriptlets level, also referred to as micro scripts, with the aim to guide a certain pedagogical concept (Dillenbourg 2002, Fischer et al. 2013). Despite these well-grounded positive effects, questions arise about whether over-scripting of learner inputs may be deficient to learning because of inhibiting autonomy that is especially needed when being confronted with open-ended argumentation learning tasks (Dillenbourg 2002, Dillenbourg et al. 2009, Scheuer et al. 2012, Noroozi et al. 2020).

When drawing upon behavioral modeling from social cognitive theory as a kernel theory and the theoretical mechanism of observational learning (Bandura 1986, Schunk 2012, Sullivan et al. 2022), we argue that vicarious experiences through an observational approach to learn patterns of argumentation might be superior for several reasons. Because previous research lacks empirical studies that compare scripted modeling (e.g., the proven upfront argumentation scripting instructions) with the outlined dynamic argumentation modeling approach, we argue that it is crucial to uncover whether rigid scripting is still superior to flexible and less structured argumentation learning approaches. However, the implementation of these ML-based systems is a complex endeavor that needs to be studied from an interdisciplinary perspective based on psychology, didactical design, HCI, and educational technology. Hence, as Scheuer (2015) and Stab and Gurevych (2017b) mentioned, current research lacks a rigorous investigation of

argumentation feedback approaches in controlled empirical studies.

By dynamically monitoring and modeling students' argumentation behavior (Scheuer et al. 2010, Stab and Gurevych 2017b, Noroozi et al. 2018, Afrin et al. 2021), the provided behavioral models should define goals, monitor progress toward the goals, and identify the activities that will enable a learner to achieve a certain goal (Hattie and Timperley 2007). Put simply, models provided through an argumentation mining algorithm could provide previously not studied sources of vicarious experiences. However, the adaptivity level can differ significantly. Hence, it is important to specifically regard the effect of different granularity levels of adaptivity to ensure that our ML-based dynamic model (meaning recommendations based on the learner's individual argumentation skill) really helps them to learn how to argue. In the vein of social cognitive theory, Bjork et al. (2013) stated that providing possibilities of feedback to evaluate one's own learning progress is a fundamental component of effective learning. Nevertheless, humans struggle to monitor and evaluate their learning and comprehension of complex learning tasks such as argumentation (Bjork et al. 2013). Hence, the design of appropriate models for the evaluation and feedback for certain skills might help learners to continuously learn more effectively (Zimmerman and Schunk 2001, Roediger and Karpicke 2006). Furthermore, research has suggested that feedback on learners' errors through alignment of one's own behavior and the dynamic argumentation model can greatly facilitate new learning (Metcalf 2017). According to Metcalf (2017), making errors and receiving feedback "enhances later memory for and generation of the correct response, facilitated active learning, [and] stimulates the learning to direct attention appropriately."

In addition, providing learners with greater control over their learning process can lead to positive impacts on learning effectiveness and outcomes, especially in more open-ended tasks such as argumentation with reasoning across multiple arguments (Brown et al. 2016). Therefore, we argue that not restricting learners upfront as in scripted learning and making additional enactive experiences as proposed in social cognitive theory are crucial for learning. In this context, Bell and Kozlowski (2008) identified the related concepts of guided exploration and error management for successful learning. Guided exploration allows learners to act on and receive feedback from a digital learning tool, whereas error management motivates educational designers to develop pedagogical scenarios in which learners can make errors and inductively discover strategies to improve and learn accordingly (Brown et al. 2016, Zacher and Frese 2018). In consequence, addressing learner errors aims to correct wrong assumptions or missing knowledge in the learner's knowledge base

to improve learning processes and, consequently, improve learning outcomes (Metcalf 2017). Accurate and detailed argumentation modeling of erroneous argumentation is most helpful and effective, as evidenced by improvements in writing outcomes (Afrin et al. 2021). Research has also found that transparent argumentation highlighting (e.g., with in-text colors) in combination with background information and explanations is more effective in helping users understand the feedback through the provided models (Afrin et al. 2021). In particular, colored in-text feedback on claims and premises has been shown to successfully support argumentation skills (Zhang et al. 2016, Chernodub et al. 2019, Afrin et al. 2021). However, dynamic models have been repeatedly mentioned as a requirement for users to successfully understand and incorporate feedback (Hattie and Timperley 2007, Afrin et al. 2021, Xu et al. 2021). In consequence, we hypothesize that despite the proven effects of scaffolding argumentation learning through scripted inputs, dynamic argumentation modeling improves persuasive writing performance through vicarious experiences of observational learning and the possibility of making an enactive experience that enables to correct erroneous argumentation.

**H1.** *Dynamic argumentation modeling improves persuasive writing performance more so than scripted argumentation modeling.*

### Dynamic Argumentation Modeling vs. Adaptive Support

As we outlined in Table 1, besides scripted argumentation modeling, there is also a plethora of adaptive support approaches for argumentation learning (see, e.g., McLaren et al. 2010 and Scheuer et al. 2014). Because these adaptive approaches also provide a valuable basis for the development of our approach based on social cognitive theory, we should further examine whether our theory-based approach is superior to basic adaptive support approaches. Adaptive support for argumentation learning involves system adjustments based on argumentation learning analysis, using adaptive strategies to enhance the learning process (Scheuer 2015). Typically, this support is provided on demand (as in our dynamic approach) and in most cases through textual feedback or simple dashboard-like charts (see the review of Scheuer 2015 but also recent examples such as Guo et al. 2023). We argue that the embedding of dynamic argumentation modeling that considers more sophisticated argumentation modeling through incorporating dynamic highlighting, graph-like representations of argumentation structures, and focusing on learners' errors is superior to only providing adaptivity through textual feedback or dashboard-like charts in combination with feedback messages based on a learner's input (Scheuer et al. 2012, Guo et al. 2023). Thus,

we further argue that adaptivity per se is not the key to providing effective argumentation learning; instead, our thoughtful embedding of adaptivity through social cognitive theory and its faithful instantiation into the user interface is more important. As outlined in the recent review of Guerraoui et al. (2023), adaptive feedback through approaches such as highlighting does not work on its own and should be accompanied by other techniques that allow learners to fully explore the dynamic modeling of an argumentative text. Thus, we hypothesize the following:

**H2.** *Dynamic argumentation modeling improves persuasive writing performance more so than adaptive support.*

### Task Difficulty in Dynamic Argumentation Modeling vs. Adaptive Support

After elaborating on two more general hypotheses related to whether dynamic argumentation modeling is superior to scripted modeling (H1) and adaptive support approaches (H2), we want to dig deeper into the theoretical mechanisms of social cognitive theory and argumentation learning when providing dynamic modeling to enable observational learning processes and positive outcomes. Because the working memory of individuals has limited capacity, a major aspect of learning is the inherent difficulty of the task (Sweller et al. 1998), that is, in our case, an argumentation learning task. Task difficulty in learning tasks is typically described through element interactivity (Marcus et al. 1996), resulting in a varying degree of complexity. In an argumentation task, we have multiple disparate elements of an argumentation that interact with each other, for example, multiple claims and premises that form arguments and counterarguments. This is especially important when writing argumentative texts because effective argumentation includes the formulation of arguments but also the proper evaluating of arguments, weighing arguments, and combining the arguments for a conclusive statement (Shehab and Nussbaum 2015). Assuming that tasks become more complex, we argue that more sophisticated and dynamic modeling enables a better comprehension of complex argumentative structures, for example, in our case, enabled through graphs displaying the structures of an argumentation and the vivid highlighting of the machine-learning model classifications directly in an argumentative text. In consequence, we suggest that dynamic argumentation modeling is superior compared with other machine-learning-based approaches that enable adaptive support for argumentation learning through textual feedback. Nonetheless, as Nussbaum (2008) highlighted, the different elements of argumentation that are also present in simple tasks impose the need to have a high degree of support. In consequence, we assume that the proposed effect of

our dynamic argumentation modeling approach is also favorable (compared with the adaptive approach) when facing more simple tasks. Hence, we hypothesize the following:

**H3.** *Compared with adaptive support, dynamic argumentation modeling leads to better persuasive writing performance on both complex and simple tasks.*

### Dynamic Argumentation Modeling vs. Static Argumentation Modeling over Time

Because prior IS research draws mainly upon static behavioral modeling, we further want to argue why behavioral modeling of argumentation on a dynamic basis is superior to providing static models of desired behavior. In the following, we ground these effects in the four underlying observational learning processes: attention, retention, production, and motivation (Bandura 1986). With static modeling, we imply an instructional method that provides non-personalized instructional feedback (e.g., knowledge of results feedback or example-based learning; for example, see van Gog and Rummel 2010, Carter and Ste-Marie 2017). This has been proven to be an effective learning intervention in technology-mediated learning across domains (van Gog and Rummel 2010, Carter and Ste-Marie 2017). We highlight that providing dynamic ML-based modeling constitutes a novel type of behavioral modeling and provides a higher degree of observational learning related to argumentation. First, attention processes are directed through accentuating relevant task features subdividing complex activities with a competent model and demonstrating the usefulness of modeled behaviors. Whereas static models such as static text- or video-based models (see for instance Gupta and Bostrom 2013, Sullivan et al. 2022) provide only predefined models for observational learning and attention processes, our dynamic ML-based modeling approach accentuates relevant aspects of the underlying task on an individual basis and highlights, after adapting behavior individually, how successful this adaption was. Second, retention of information is increased through rehearsal and coding of information in visual and symbolic form. In contrast to static approaches such as symbolic coding through notetaking and rehearsal such as in Yi and Davis (2003), our approach fosters retention through a deep connection of learning with own behavior. As the argumentation mining algorithm highlights argumentative structures, cognitive organizing is facilitated, and rehearsal is embedded with the own practicing, thus contributing to a better retention of argumentation skills. Third, during production processes, behavior is compared by learners to their own conceptual (mental) representation. Whereas static models provide feedback that is generic (e.g., knowledge of results feedback or example-based learning; for example, see van Gog

and Rummel 2010, Carter and Ste-Marie 2017), dynamic AI-based models provide feedback that helps to correct deficiencies through more situated feedback that is adapted to the actual behavior. Fourth, and finally, motivational learning processes are important for observational learning because this determines the engagement with the aforementioned three processes (Schunk 2012). Whereas static models provide motivation only on a generic level, for example, if a learner recognizes that a model helped to show a certain behavior, dynamic ML-based modeling tailors the consequences of modeled behavior to the actual learner behavior and informs precisely about outcome expectations.

In summary, dynamic and ML-based learning processes provide a novel approach to behavioral modeling and higher degrees of observational learning compared with approaches that are prevalent in IS research. On this basis and mechanisms described above, we hypothesize that this type of learning process mechanism can help students develop argumentation skills by observing and drawing upon the tailored feedback provided by the AI system. Therefore, we move beyond existing approaches in IS research that are drawing mainly upon static behavioral modeling approaches. In addition, there exist no studies that investigate the impact of dynamic argumentation feedback in a long-term field experimental setup. Especially previous research on argumentation feedback has often neglected the impact of argumentation learning tools on argumentation skills—meaning the impact of dynamic argumentation feedback in one pedagogical domain (e.g., business argumentation) on general argumentation skills in other domains (e.g., social debates). Hence, we suggest the following hypothesis to investigate the impact of dynamic argumentation modeling on argumentation skills in a static approach:

**H4.** *Dynamic argumentation modeling on repeated argumentation tasks (over three months) leads to better learning in comparison with static argumentation modeling.*

## Experimental Studies

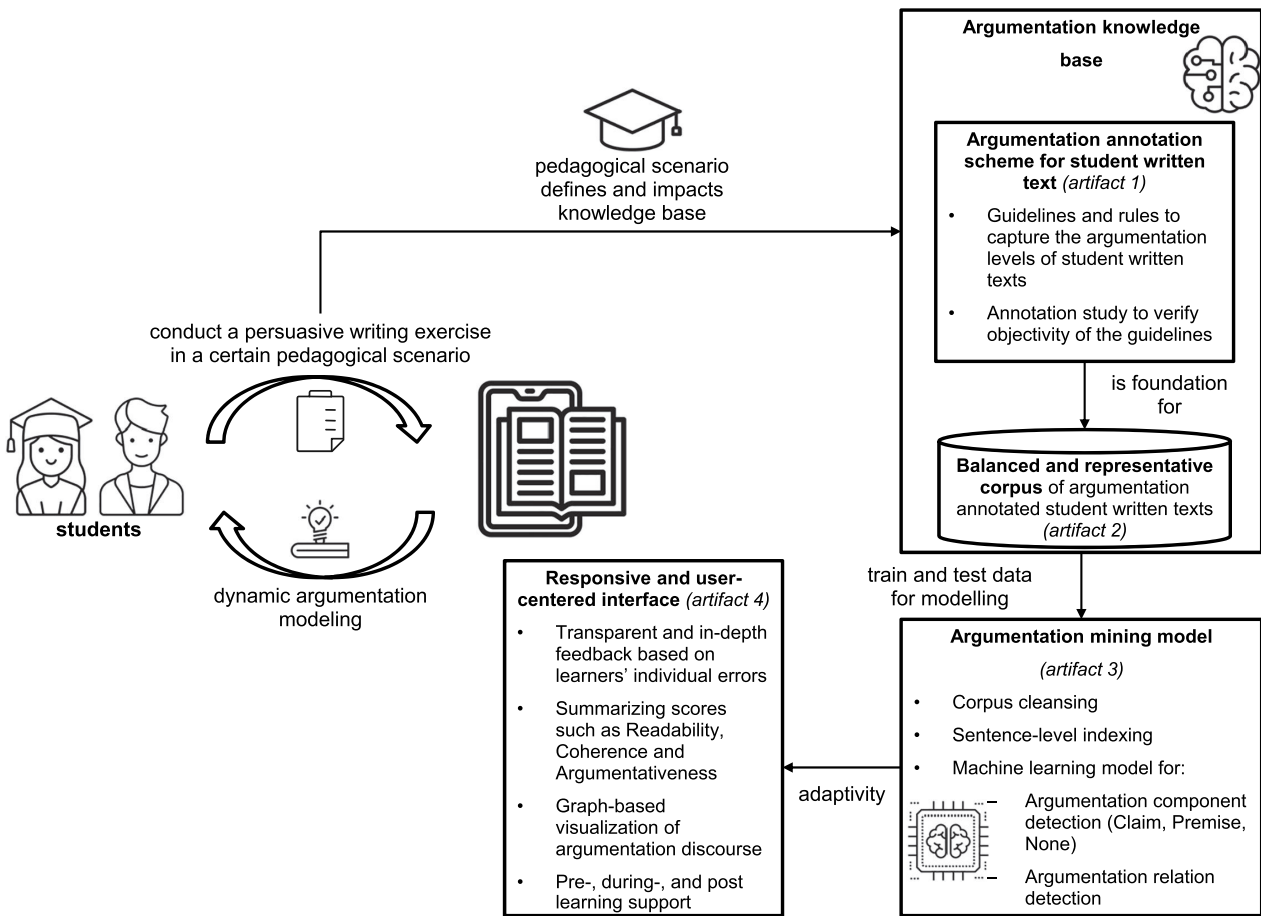
To test our four hypotheses, we designed three experiments to manipulate the argumentation modeling approaches (i.e., dynamic argumentation modeling and scripted modeling for study 1; dynamic argumentation modeling and adaptive support as well as task difficulty for study 2; dynamic argumentation modeling, static argumentation modeling, and no modeling for study 3). The dependent variables in study 1 and study 2 are the objective and the subjective quality of argumentation of students' text according to the scheme of Weinberger and Fischer (2006). In study 3, we investigated the impact of the argumentation modeling approaches on the argumentation skill of students in an argumentation

task outside of the domain of the course content (Weinberger and Fischer 2006).

## Implementation of a Theory-Driven Dynamic Argumentation Modeling System Based on ML

For this study, we developed an argumentation modeling system based on ML and NLP. Similar to our hypotheses, we informed the design based on theory about (argumentation) skill learning (see, e.g., Pinkwart et al. 2009, Jonassen and Kim 2010, Wang et al. 2020), argumentation theories (Kuhn 1993, Toulmin 2003, Walton et al. 2008), technology-mediated learning (Alavi and Leidner 2001, Hattie and Timperley 2007, Gupta and Bostrom 2009, Han et al. 2021), literature on argumentation mining, and the design of educational learning tools (see, e.g., Mayer 2014, Santhanam et al. 2016).

For the system class of dynamic argumentation modeling approaches, we propose a service-oriented architecture approach (Niknejad et al. 2020, Slonim et al. 2021). A service-oriented architecture is composed of multiple modules with a set of defined functionalities. In our research, we refer to the modules as artifacts, following the design science research paradigm (Gregor and Hevner 2013). The artifacts can operate and be updated independently. Thus, a service-oriented architecture allows for high availability and interoperability for different domains, use-cases, and user groups (Niknejad et al. 2020). This fosters the reproducibility of our artifacts for adoption to other use cases (see, e.g., for other languages or learning tasks). The system consists of four artifacts; the core of the architecture is the pedagogical scenario, in which learners conduct a persuasive writing exercise and receive ongoing and individual argumentation evaluation through dynamic argumentation modeling irrespective of instructor, time, and location (see Figure 2). The pedagogical scenario defines and impacts the argumentation knowledge base. The argumentation knowledge base comprises an argumentation annotation scheme for students' written text with guidelines and rules to capture the students' argumentation levels in the particular pedagogical scenario (*artifact 1*). The evaluated annotation scheme is the foundation of an argumentation annotated corpus of train and test data (*artifact 2*) for the argumentation mining algorithm. The argumentation mining module (*artifact 3*) is trained on the corpus, utilizing a machine-learning approach to access the individual argumentation level in student-written texts.<sup>1</sup> The advantage of machine learning compared with argumentation skill feedback in the literature (see, e.g., Scheuer et al. 2010, Lin et al. 2015, Huang et al. 2016) arises from its broader applicability, scalability, and capacity to improve based on ongoing usage through the generation of new training data and self-learning capabilities (Lawrence and Reed 2019).

**Figure 2.** Overview of a Dynamic Argumentation Modeling System's Architecture Based on ML

Finally, the responsive and student-centered interface enriches the students' evaluation with individualized, transparent, and in-depth skill feedback (artifact 4).

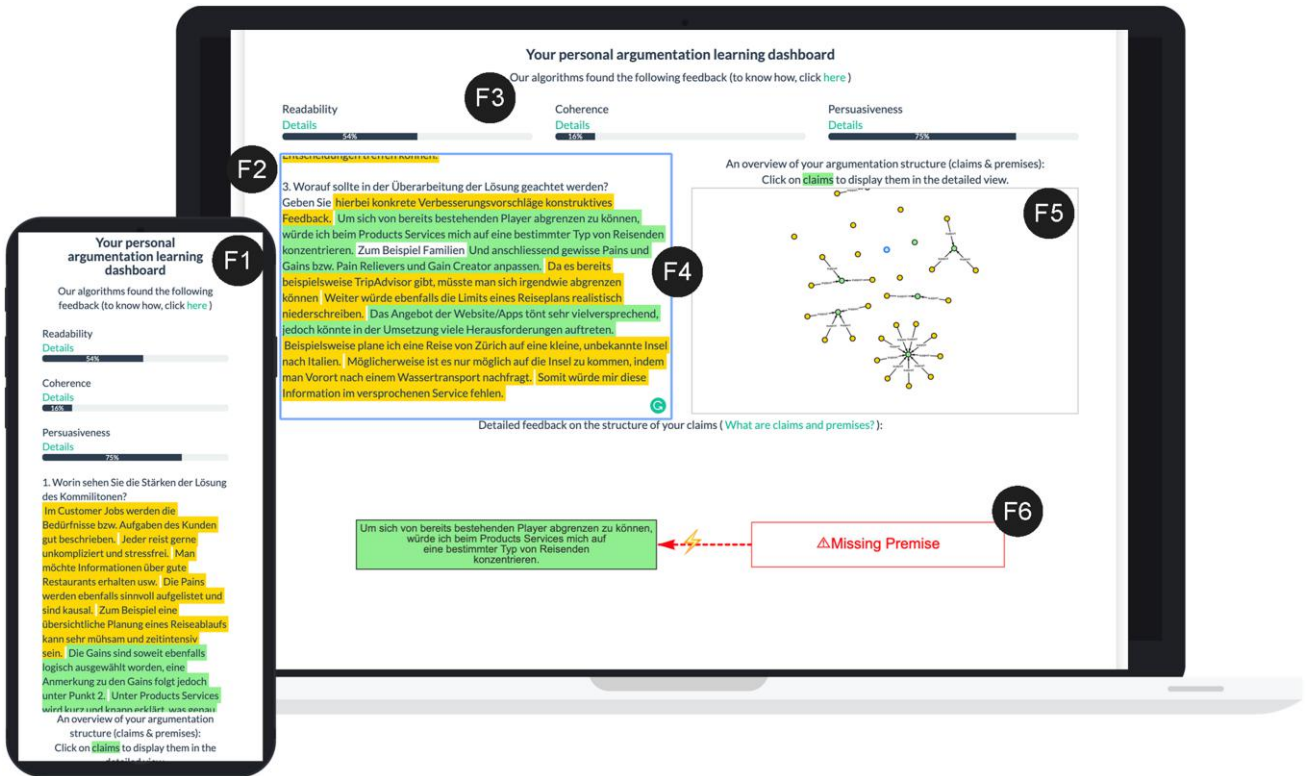
The specific instantiation was informed by 30 semi-structured interviews with students to form a concise approach and learning system (Hevner et al. 2004). The interviews eventually enriched the design with knowledge from the field and the end users and helped us build a student-centered learning tool.<sup>2</sup> First versions of our artifacts from this three-year interdisciplinary research project have been partly published at different conferences in several domains, for example, the user-interaction artifact in HCI (Wambsganss et al. 2020a), the socio-technical system design concept in IS (Wambsganss and Rietsche 2019), or the argumentation annotation scheme, the corpus, and the predictive model in NLP research (Wambsganss et al. 2020b). Our final instantiation *ArgueLearn* allows users to input text and receive dynamic modeling on their argumentation. A screenshot of *ArgueLearn* with exemplars of functions (see, e.g., F1–F6) is presented in Figure 3.

### User Interface Design

We built *ArgueLearn* as a responsive Web-based application for use on various devices. The front end of *ArgueLearn* was developed with recent Web technologies, including HTML5, Cascading Style Sheets (CSS), and JavaScript (JS).

A user can access *ArgueLearn* from any Web-enabled device regardless of screen size or operation system and can log in with a single sign-on for easy access (F1). It provides the user with a simple and intuitive text input field and a word count (F2) in which they can write or copy a text. Next to the input field, the users receive feedback on their text's argumentation structure on a personal learning dashboard (F3–F6). By this means, the feedback resembles the adaptive modeling of the desired learner behavior and according to observational learning opportunities. The dashboard provides feedback on different granularity levels, allowing users to control the amount of feedback information required. A visual graph-based representation of the submitted text's argumentation structure (F5) and three summarizing scores give an initial overview of the

**Figure 3.** (Color online) Screenshot of Our System *ArgueLearn* Providing Dynamic Modeling on Students' Argumentation Skills with Instantiated Functionalities (F1 to F6)



text's quality (F3). The identified claims are green, and the premises are highlighted in yellow to provide users with instant feedback on their own submitted input (F4). By clicking on the marked text fields or the nodes in the graph, a more detailed view of the discourse of the argument will appear (F6). This indicates whether a claim is sufficiently supported (see Figure 3) or if it lacks a premise (as in F6). This function provides learners with clear steps on how to improve their texts' persuasiveness and formal quality. Moreover, best practices and explanations about argumentation and argumentation theory are provided by clicking the "explanation" or "help" button (F7), providing the user with an orientation and context to improve their writing quality. The three summarizing scores—readability, coherence, and persuasiveness (F3)—provide the users with a ranking of their text for superficial instant feedback. By clicking on the scores or on "details," the methodology for calculating the scores, as well as concrete hints and explanations of how the learner can increase their score level, will be shown (F7). The scores are calculated using the following metrics; readability is defined as how readable the text is based on the Flesch reading ease score (Flesch 1943); coherence indicates the proportion of sentences that are connected via discourse markers; and persuasiveness concerns the proportion of claims that are

supported by premises compared with unsupported claims. F6 and F7 are not visible in Figure 3 but were implemented on different screens.

### Dynamic Argumentation Modeling Based on NLP and ML

In order to model learners' argumentation skills and to provide adaptive feedback, high-quality annotated corpora are necessary for training certain ML models to learn and predict behavioral patterns. The basis for the annotated corpora is evaluated annotation schemes. Because no suitable argumentation annotation was available to guide the annotators to a substantial agreement for persuasive student essays in the German language, we decided to create a new annotation scheme (artifact 1) for the pedagogical scenario of writing peer reviews and conducted a corresponding annotation study. Thereby, we presented the first argumentation annotation scheme for student peer reviews in the literature and the first annotation scheme for persuasive student texts in a language other than English. We derived an annotation scheme for a new data domain for AM based on argumentation theory and previous work on annotation schemes for persuasive student essays (see, e.g., Stab and Gurevych 2014). Our objective was to model the argumentation discourse structures of

student-generated peer reviews by annotating the argumentation components and their relations. We chose student peer reviews because they form a modern, scalable, growing, and domain-independent pedagogical scenario that can be adopted in a scalable way to tutor students' argumentation across domains. The first step in creating an annotation scheme is to model theoretical structures in the text domain. Therefore, we collected a new data set of student-written peer reviews. The data were collected in a mandatory business innovation lecture in a master's program at a Western European university: around 220 students developed and presented a new business model for which they received three peer reviews in which a student from the same course elaborated on their model's strengths and weaknesses and gave persuasive recommendations on what might be improved. We collected around 7,000 documents from 2012 to 2019.

Based on our novel annotation scheme (artifact 1), we randomly collected a balanced and representative set of 1,000 student-generated peer reviews written from our lecture. The corpus was evaluated in prior research efforts (Wambsganss et al. 2020b). To build a reliable corpus, we followed a four-step methodology, following the established argumentation annotation process provided by (Stab and Gurevych 2017b). (1) We examined scientific literature and theory on how to model argumentation discourse structures in texts from different domains; (2) we randomly sampled 50 student-generated peer reviews and, based on our findings from literature and theory, developed a set of annotation guidelines comprising rules and limitations on how to annotate argumentation discourse structures (artifact 1); (3) we applied, evaluated, and improved our guidelines with three native-speakers in three consecutive workshops to resolve annotation ambiguities; and (4) we applied the final annotation scheme based on our 15-page guidelines to a corpus of 1,000 student-generated peer reviews.<sup>3</sup> The final corpus consists of 1,000 student-written peer reviews in German, amounting to 20,125 sentences with 246,980 tokens in total. On average, each document has 20 sentences and 272 tokens. A total of 7,996 claims (31.64%) and 8,479 premises (33.55%) were annotated, and 8,796 textual spans (34.81%) were not identified as argument components.

To dynamically model learners' argumentation, we implemented an approach to identify arguments in their texts. This approach comprises two subtasks. First, we identified the arguments' components in terms of claims and premises. Next, we determined whether an argumentative relationship existed between a pair of components, following the approach described by Stab and Gurevych (2017b) to identify argumentative discourse structures in persuasive essays. The identification of argument components is considered a sentence-level multiclass classification task, whereby each sentence in

the data set is labeled as either claim, premise, or nonargumentative. To ensure an equal distribution of classes among training and test sets, we performed a stratified split of the data set into an 80% training set and a 20% test set, resulting in the distribution of 32% claims, 32% premises, and 36% nonargumentative spans (for both training and test sets). We evaluated and tuned different models along a variety of handcrafted text features in several technical experiments. We found that a Support Vector Machine (SVM) achieved the best results, with an accuracy of 65.4% on the test set. The identification of argumentative relationships is considered a binary classification task in which each argument component pair is classified as either support or non-support. All possible combinations were tested. After several technical evaluations, we found that an SVM achieved the best results for our corpus, obtaining an accuracy of 72.1% on the test set. This is satisfying compared with other studies on student-written argumentation identification. For example, (Stab and Gurevych 2017a) reached an f1-score of 73% for argumentation stance classification of student-written texts in English. In multiple technical experiments, we used several classifiers (SVM, Logistic Regression, Random Forest, Multinomial I Bayes, Gaussian NB, Nearest Neighbor, and AdaBoosted Decision Tree), and we feature combinations for the task of argument component identification and for argumentation relation classification. To tune our models' parameters, we applied grid search. We used several preprocessing pipelines, extracted several lexical and syntactical features, and iteratively tested them in evaluation cycles to tune the accuracy, precision, and recall (Wambsganss et al. 2020b). We benchmarked our final feature-based SVM model for argumentation component detection against a transformer-based Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al. 2018) and a bidirectional Long-Short-Term-Memory-Conditional-Random-Fields classifier (BiLSTM-CRF). In combination with the corresponding embeddings vocabulary (GloVe), our LSTM only reached an unsatisfying f1 score of 57%. The BERT model reached a macro f1 score of 73% for classifying text tokens into claim, premise, or nonargumentative tokens. However, we compared the actual prediction of the BERT model in our pilot study against the predictions of our feature-based SVM. We found the SVM to be more robust and reliable on unseen student-written data. Hence, we choose the more robust feature-based SVM for the final instantiation of our dynamic argumentation modeling tool.

## Experimental Design

In this section, we describe the investigation of our two hypotheses based on the application and evaluation of the theory-driven dynamic argumentation modeling system in the context of higher education. We present

one pilot study, one eye-tracking study, and three experimental studies (study 1, study 2, and study 3) with a total of 366 students.<sup>4</sup>

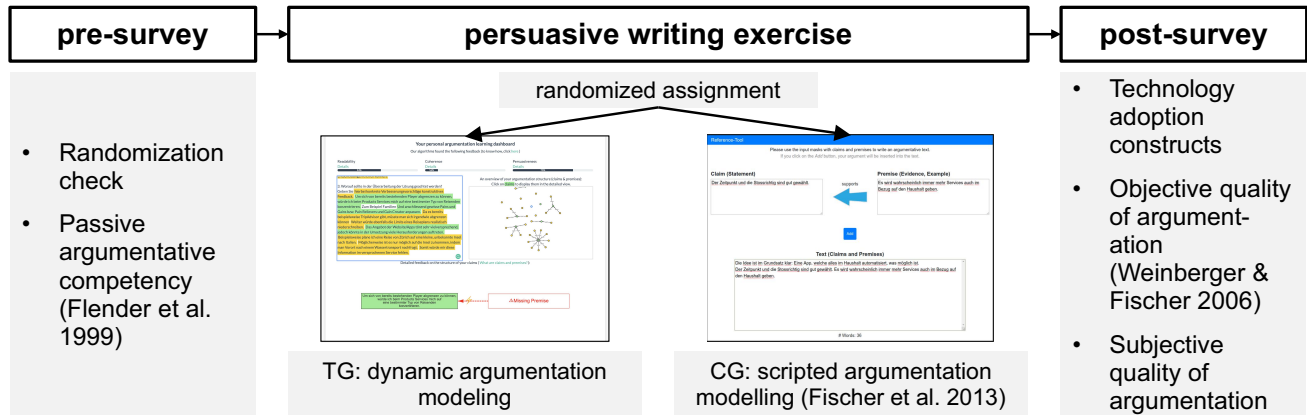
We performed a controlled laboratory experiment with *ArgueLearn* in study 1 to test *whether dynamic argumentation modeling improves persuasive writing performance more so than scripted argumentation modeling* (H1). For testing H2 (*comparison of dynamic modeling with adaptive support*) and H3 (*comparison of complex and simple task*), we employed a fully randomized 2 (argumentation learning tool: dynamic modeling versus adaptive support) × 2 (task difficulty: complex argumentation task versus simple argumentation task) between-subjects design. To do so, we conducted the exact same controlled laboratory experimental design as in study 1. Finally, we conducted one long-term field experiment to investigate *whether dynamic argumentation modeling on repeated argumentation tasks (over three months) leads to better learning in comparison with static argumentation modeling* (H4). Study 1 and study 2 were conducted in a university laboratory designed for behavioral studies under controlled circumstances. Study 3 was conducted in a large-scale master's lecture at our university. Thus, we could ensure that no participant took part in multiple studies. Before the three empirical studies, we performed several small-scale evaluations of our distinct artifacts. Our aim was to conduct iterative evaluations to ensure that the instantiated design functionalities correctly addressed their purpose and that the technical artifacts (artifact 1, 2, and 3) enabled the learners to receive individual evaluations irrespective of instructor, time, and place. For our pilot study, we designed clickable mockups of our dynamic argumentation modeling system. For the evaluation, we followed an *ex ante evaluation* of the developed artifact using an artificial evaluation setup, as proposed by Venable et al. (2016). Based on (Venable et al. 2012) criteria, the pilot study design evaluation's objective was to (1) verify the utility and value of the artifact to achieve its stated purpose, (2) identify weaknesses and areas for improvement in the artifact design, and (3) identify change requests from students, side effects, or undesirable consequences of its use.

To reach this goal, we conducted four evaluation series with 46 different users (around 12 users per series). We performed different qualitative assessments involving observation, participant feedback, and unstructured interviews (Tuunanen and Peffers 2018). These users differed from those recruited for studies 1 and 2 as well as those from the semi-structured interviews but were also students from our university with a similar age and gender distribution. Following the pilot study design evaluation, we identified several requirements for the overall design of a learner-centric dynamic modeling system for argumentation skills. We observed that learners aimed to receive more transparent explanations about the argumentative writing

goal, theory embedding, and a clear learning goal that guided them before, during, and after the argumentative writing task. Accordingly, we implemented a specific goal, purpose, and orientation in our feedback system to help learners reflect and provide guidance on the context and task (Soloway et al. 1994). Moreover, we conducted 13 eye-tracking studies with the iterated prototype to specifically evaluate and observe how students interact with erroneously predicted argumentation modeling structures that may impact learning outcomes according to prior research (see, e.g., Schmitt et al. 2021). We saw that 10 of the 13 students took erroneous highlighting into account but would not change their argumentation if the model predicted clearly wrong argumentation components or relations. In a qualitative interview conducted after the eye-tracking study, 9 out of 13 learners noted that the design of *ArgueLearn*, which highlights argumentative errors without actively suggesting corrections, ensures that erroneous modeling predictions are not overly emphasized. Still, we introduced *ArgueLearn* to students as a “trained” system (“also a student like you, that is learning and possibly sometimes making unintended mistakes”) in the subsequent studies to provide indications of possibly erroneous predictions.

### Study 1: Evaluation of Dynamic Argumentation Modeling vs. Scripting in Laboratory Experiment

After evaluating the technical artifacts and the design in a pilot study, our aim was to test whether dynamic argumentation modeling improves persuasive writing more so than scripted argumentation modeling. Hence, we conducted study 1 as a laboratory experiment in which participants were asked to provide peer feedback on an essay (see Figure 4). The treatment group used our dynamic modeling system, whereas participants in the control group used the discussion scripting application, in accordance with Fischer et al. (2013) as a well-cited and empirically proven benchmark to foster the formal quality of argumentation. We recruited students from our university through social networks and mailing lists to participate. After randomization, we had 24 participants in the treatment group and 30 in the control group. We invited them to the laboratory of our university, where we conducted the study on identical devices. Participants in the treatment group had an average age of 23.8 (SD = 3.86); 15 were male, and nine were female. In the control group, the participants' average age was 23.03 (SD = 2.12); 22 were male, and eight were female. All participants were compensated with an equivalent of about \$15 for the 30- to 40-minute experiment. The experiment included three main phases: (1) the presurvey phase, (2) the individual writing phase, and (3) the postsurvey phase. The pre- and postsurvey phases were the same for all participants. In the writing phase, the

**Figure 4.** (Color online) Overview of Experimental Procedure of Study 1 to Evaluate Whether Dynamic Argumentation Modeling Improves Persuasive Writing More So Than Scripted Argumentation Modeling

treatment group used *ArgueLearn*, and the control group used the alternative tool developed by Fischer et al. (2013).

**1. Presurvey Phase.** The experiment began with a pre-survey comprising 14 questions. Here, we tested three different constructs to assess whether randomization resulted in comparable groups. First, we had four items to test the participants' personal innovativeness in IT, following Agarwal and Karahanna (2000). Next, we tested individuals' feedback-seeking following Ashford (1986). Example items included, "It is important for me to receive feedback on my performance," or "I find feedback on my performance useful." Both constructs were measured using a five-point Likert scale (1: totally agree to 5: totally disagree; 3: neutral). Third, we captured the construct of passive argumentative competency, following the design of Flender et al. (1999), because it is a proven construct to measure argumentative competencies. We wanted to control for argumentative competencies because we later measured the objective and subjective quality of the texts' argumentation. Participants were asked to read a discussion between two teachers concerning the following topic: "Does TV make students aggressive?". We retrieved the topic with the discussion as well as the measurements along with nine questions from Flender et al. (1999).

**2. Writing Phase.** During the writing phase, we asked the participants to write a review about both parties' argumentation (pro and contra) with respect to the weaknesses and strengths. The participants were told to spend at least 15 minutes writing this review, with a countdown indicating the remaining time. They were able to continue the experiment only after the countdown had finished. The treatment group used *ArgueLearn* to write the review, and the control group used the scripted argumentation modeling tool. We provided

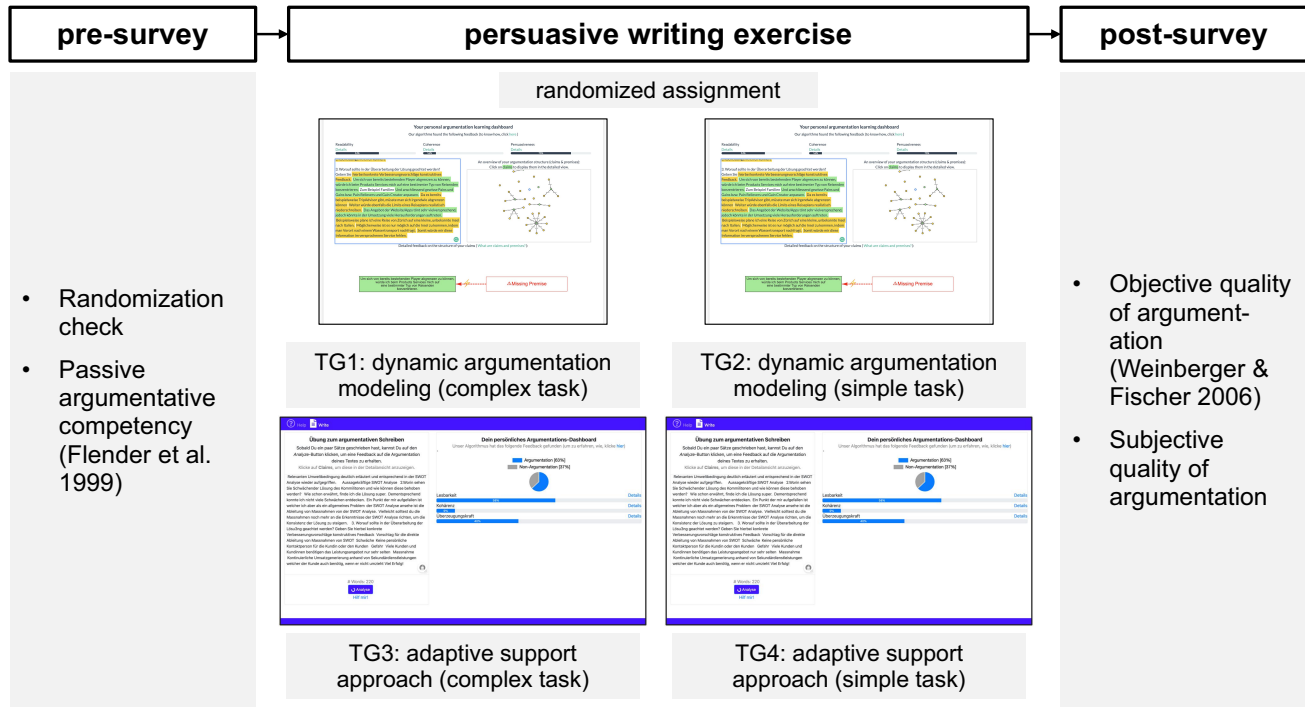
no introductions to any of the tools. The students using *ArgueLearn* received adaptive support based on dynamic argumentation modeling. Participants in the control group received help based on scripted input formats during the writing process. Both approaches (dynamic ML-based modeling and the scripted input) are based on the argumentation quality model of Toulmin (2003) (i.e., Figure 1). In the treatment group, students wrote their text first, without seeing the initial feedback. Only after clicking the analyze button (if wished repeatedly after their revisions), they would receive adaptive feedback based on our dynamic ML-based modeling. For the control group, students were presented with an input mask for writing their argumentation right from the beginning of the exercise. Similar to Fischer et al. (2013), the input script was always present and guided students in their writing. Nevertheless, instead of writing into the claim and premise fields, they could also enter their argumentation directly into the text box as done in the approach of Fischer et al. (2013).

**3. Postsurvey Phase.** In the postsurvey, we measured perceived usefulness, intention to use, and ease of use, following the technology acceptance model (Venkatesh et al. 2003, Venkatesh and Bala 2008), and captured the demographics. Moreover, we asked three qualitative questions: "What did you particularly like about the use of the argumentation tool?"; "What else could be improved?"; and "Do you have any other ideas?" to control for boundary conditions and qualitative effects (e.g., erroneous modeling).

### Study 2: Evaluation of Dynamic vs. Adaptive Argumentation Modeling in a Laboratory Experiment

In study 2, we aimed to test whether *dynamic argumentation modeling improves persuasive writing performance more so than adaptive support (H2) and whether, compared*

**Figure 5.** (Color online) Overview of the 2 × 2 Experimental Procedure of Study 2 to Evaluate Whether Dynamic Argumentation Modeling, in Comparison with Adaptive Support Approaches, Leads to Better Argumentation



with adaptive support, dynamic argumentation modeling leads to better argumentation on both complex and simple tasks (H3). To dive deeper into task difficulty and the comparison of dynamic modeling and adaptive support, we aimed to investigate those two hypotheses together in one controlled laboratory experiment.

The experiment was conducted in the exact same procedure as study 1 (including the same compensation, the same presurvey phase, and the same postsurvey phase). We changed only the writing phase of the experiment. We manipulated (1) the *argumentation learning tool learners used for the task* (dynamic modeling versus adaptive support), and (2) the *task difficulty* (complex argumentation task versus simple argumentation task) in a between-subjects design, resulting in four treatment groups (TG1–TG4; see Figure 5). Our objective was to investigate the impact

of dynamic argumentation modeling in comparison with traditional adaptive support approaches on students’ objective and subjective quality of argumentation. Moreover, we aimed to investigate the impact of dynamic modeling on students’ formal and subjective argumentation skills when conducting a complex argumentative task versus a simple argumentative task.

Again, we recruited students from our university through social networks and mailing lists to participate. After randomization, we counted 30 participants in treatment group 1 (TG1), 34 participants in TG2, 41 in TG3, and 37 in TG4 (see Table 2 for demographics). As in study 1, participants were compensated with an equivalent of about \$15 for a 30- to 40-minute experiment.

In the writing phase of the experiment, we manipulated (1) the support students received to conduct the

**Table 2.** Overview of the Treatment Groups and Their Demographics in Study 2

| Group | Argumentation support | Task         | N   | Age                       | Gender                 |
|-------|-----------------------|--------------|-----|---------------------------|------------------------|
| TG1   | Dynamic modeling      | Complex task | 30  | Mean = 24.00<br>SD = 3.30 | 19 males<br>11 females |
| TG2   | Dynamic modeling      | Simple task  | 34  | Mean = 23.34<br>SD = 4.46 | 20 males<br>14 females |
| TG3   | Adaptive support      | Complex task | 41  | Mean = 23.03<br>SD = 4.68 | 26 males<br>15 females |
| TG4   | Adaptive support      | Simple task  | 37  | Mean = 23.78<br>SD = 5.04 | 24 males<br>13 females |
| Total |                       |              | 142 |                           |                        |

argumentative writing tasks and (2) the difficulty of the task. Students in TG1 and TG2 were using *ArgueLearn*, receiving dynamic argumentation modeling during their argumentative writing task. TG3 and TG4 used an adaptive support approach, following the works of McLaren et al. (2010) or Guo et al. (2023), receiving support through adaptive pre-scripted feedback messages with improvement suggestions and through dashboard-provided meters on *readability*, *coherence*, and *persuasiveness*. The adaptive support approach was natively developed to control the differences and similarities in the manipulations. The meter-provided scores were based on the two same ML-based models as in *ArgueLearn* and based on the results calculated on heuristics (see Figure 3, F3). The writing field, the word count, the instructions, and back end were the same between all four groups, and we manipulated only the difference between traditional adaptive support approaches (with help messages and meters) and dynamic modeling based on social-cognitive theory.

For the task difficulty, we followed the approach of Shehab and Nussbaum (2015) to adjust the task difficulty of the argumentative writing task students conducted. As a basis for manipulating the task difficulty of students in argumentative writing tasks, we used the task design from study 1, because it was based on the construct of passive argumentative competencies of Flender et al. (1999) and has proven to rigorously guide students of our sample group through an argumentative exercise. To manipulate the task difficulty of the very task from study 1, two different options existed: (1) Adjust the information input of the argumentation that students must write about (the discussion of whether TV makes one aggressive retrieved from Flender et al. (1999)) or (2) adjust the task description and objective for students. To not influence our dependent variable and outcome measure of the objective and subjective argumentation quality according to Weinberger and Fischer (2006), we aimed to not semantically change the context of the exercise. Hence, together with one senior researcher from the field of educational psychology and one senior researcher from the area of cognitive science and educational design, we decided to (1) adjust the syntax of the discussion text (easier words, shorter, less complicated sentences, and grammar, easier to grasp) and (2) simplify the wording of the task without adjusting its goal and content. By these means, we reduced element interactivity for the task to enable easier processing of argumentative structures. We decided to take the unchanged task from study 1 as the complex argumentative task because it is based on a psychological test and comes with a certain complexity. We used the adjusted task (including an easier syntax of text discussion, including a sort of scaffolded content summarization, and a simplified but not semantically changed task) as the simple task design. Both tasks can

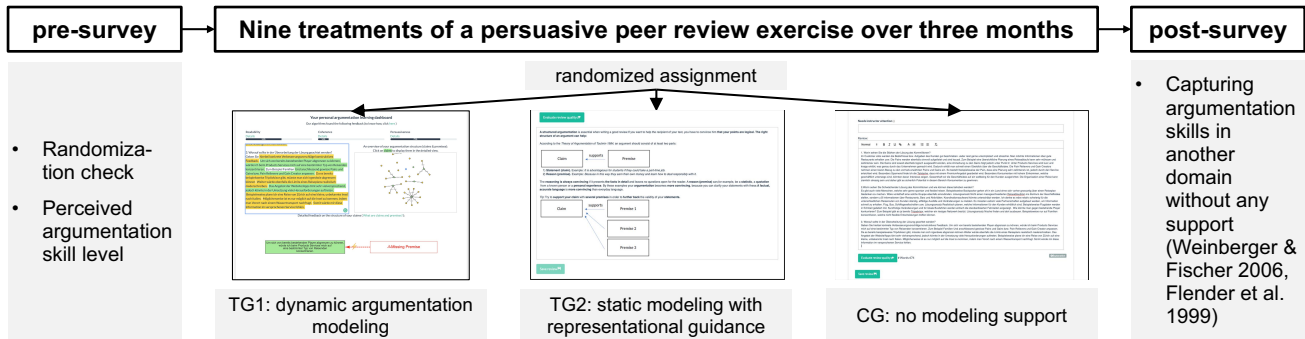
be found in the Online Appendix. To validate the task adjustments done with the two experts, we recruited five independent experts in the field of argumentative writing to judge both tasks according to the task difficulty construct of Gupta and Bostrom (2013) (with the four items: “I found this to be a complex task,” “This task might be mentally demanding,” “This task might require a lot of thought and problem solving,” and “I found this task to be challenging”) based on a 1–5 Likert Scale (1: easy; 5: difficult). The experts rated the “complex task” with a mean of 4.5 and the “simple task” with a mean of 2.5. The rating was done and is valid only for our target sample group of German business students. With these results, hence, we embedded the two tasks as a manipulation and conducted the experiment.

### Study 3: Evaluation of Dynamic vs. Static Argumentation Modeling in a Field Experiment

In study 3, our objective was to dive deeper into the mechanisms of argumentation modeling and whether dynamic argumentation modeling on learners’ individual errors, in comparison with static argumentation modeling, leads to better learning of argumentation skills. Our goal was to measure the extent to which the dynamic modeling of our proposed approach influences learners’ long-term argumentation skills in a large-scale learning setting (see Figure 6). Therefore, we evaluated our dynamic modeling system in a real learning environment with 205 students in the context of higher education, from which 124 participated in our surveys. We implemented *ArgueLearn* in a business innovation lecture in a master’s program at our university.

Following the Bauman and Tuzhilin (2018) design, we conducted a field experiment with three treatment groups to evaluate the impact of the ML-based dynamic modeling (provided by our algorithms) on humans’ argumentation quality. The functionalities necessary for the treatments were implemented in our learning system. In the lecture in which we conducted the field experiment, students developed and presented new business models. The students were asked to submit three assignments developing their own business idea over three months. Students were also required to peer review fellow students’ assignments (three peer reviews per student per assignment). Because the course included three rounds of assignments, each student had to write nine peer reviews to elaborate on the strengths and weaknesses of their fellow students’ business models and offer persuasive recommendations on what could be improved. The quality of the peer review (e.g., the persuasiveness) did not influence the students’ final grades. However, submitting a peer review was mandatory for completing the assignments. For peer reviews, we provided the students with different levels of argumentative

**Figure 6.** (Color online) Overview of Experimental Procedure of Study 3 to Evaluate Whether Dynamic Argumentation Modeling, in Comparison with Static Argumentation Modeling, Leads to Better Learning of Argumentation



support. Our field experiment comprised three main phases: (1) a presurvey phase, (2) an individual writing phase, and (3) a postsurvey phase. The pre- and posttest phases were consistent for all participants. In the writing phase, we manipulated the level of argumentation that feedback participants received while writing their peer reviews.

**1. Presurvey Phase.** Before the assignment period, we conducted a mandatory postsurvey to control for randomization. We checked the same items as in study 1 (i.e., personal innovativeness in the domain of information technology (Agarwal and Karahanna 2000)) and individuals' feedback seeking (Ashford 1986). Third, because no elaborated argumentation test was possible in our field setting, we asked the participants to provide an individual judgment of their argumentation skills, following the Toulmin model (Toulmin 2003), and self-determination of competencies (Vansteenkiste and Ryan 2013) using three items: "How do you rate your ability to argue?"; "How would you rate your ability to convince others?"; and "How would you rate your ability to write argumentative texts?" We wanted to control for argumentative competencies because we later measured the texts' objective quality of argumentation. These items were measured using a seven-point Likert scale (1: extremely poor to 7: excellent; 4: fair).

**2. Individual Writing Phase.** In the treatment phase, participants were asked to write nine peer reviews over three months. Following the experimental design of Bauman and Tuzhilin (2018), the first treatment group (TG1) received ML-based dynamic argumentation modeling by *ArgueLearn*; following our design, the second treatment group (TG2) received static (non-ML-based) argumentation modeling in the form of a text and a visualization of a theoretical argumentation model (Toulmin 2003). The control group (CG) received no modeling at all during the writing process. Before the writing phase, all students were presented

with the same instructions about the importance of argumentation in business model peer reviews according to the Toulmin model (Toulmin 2003). This is in line with prior field research on technology-mediated learning interventions (Gupta and Bostrom 2013). To avoid any confounds in the experimental setup, all three groups started the writing process in the same fashion, and they were presented with a text editor to write their feedback and received their writing instructions, always visible on top. Both treatment groups (TG1 and TG2) were shown an additional analyze button at the bottom of the screen, allowing them to receive dynamic or static modeling feedback (on a second screen) at any time if they so wished. Again, initially, participants of all three groups would not see any feedback, dashboard, or help. Only after clicking on the analyze button (also repeatedly possible after iterative revisions), TG1 received dynamic argumentation modeling and TG2 received static argumentation modeling. Both feedbacks are based on the Toulmin model and do not differentiate from a theoretical perspective related to the quality dimension (Toulmin 2003). In addition, all three groups were seeing the instructions during the writing. In total, 205 students were in our class and randomly assigned to one of the three groups. After three months, 124 students completed all three assignments. Table 8 presents an overview of the demographics.

**3. Postsurvey Phase.** After the treatments, we asked the students to conduct a posttest via a Web survey. Our main objective was to measure the learners' argumentation skills to evaluate our research question for study 2. Therefore, we tested whether the three groups' argumentation skills differed in a second argumentation domain. This was necessary to control whether dynamic modeling significantly impacted the argumentation skills or only humans' skills of writing more persuasive peer reviews (Toulmin 2003). Thus, we asked the participants to read the same discussion as in study 1 between two teachers concerning the following

topic: “Does TV make students aggressive?” We retrieved the topic with the discussion as well as the task from the construct of passive argumentative competency from Flender et al. (1999) because it is proven for measuring argumentative competencies. Like study 1, we asked the students to review both parties’ argumentation (pro and contra) with respect to the weaknesses and strengths. The participants were told to spend at least 15 minutes writing this review. We captured the demographics and asked three qualitative questions—“What did you particularly like about the use of the argumentation tool?”; “What else could be improved?”; and “Do you have any other ideas?”—to further evaluate the instantiations and to check for the completeness of our proposed solution (e.g., again for the effect of erroneous modeling based on our algorithm).

### Measurement of Persuasive Writing Performance

Besides measuring the technology acceptance to check for a successful instantiation of *ArgueLearn*, our main objective was to measure the persuasive writing performance of learners to test our hypotheses H1–H4. To measure the persuasive writing performance, we rely on prior literature (see, e.g., Weinberger and Fischer 2006, Stegmann et al. 2012) and the theory of argumentation of Toulmin (1984) to assess the objective quality of argumentation of learners, measured in the coherence to formal argumentation structure. Moreover, next to the objective quality of argumentation, we also aimed to assess the subjective quality of argumentation—assessed through external ratings of learners’ subjective persuasiveness of their argumentative texts. Although the objective quality of arguments on a micro level measures the formal nuances of an argument, on a document macro level, a higher number of qualified arguments also represent an overall more persuasive and high-quality text. Still, through measuring the subjective quality of argumentation, we aimed to shed extra light on the subjective persuasiveness of a learners’ argumentation (besides the formal character). Therefore, we measured two main variables: (1) the objective quality of argumentation and (2) the subjective quality of argumentation.

**1. Objective Quality of Argumentation.** The written peer reviews were analyzed for the objective quality of argumentation measured by assessing the formal structure of the arguments, for example, if argumentative claims are sufficiently bagged up by premises. As previously explained, according to formal argumentation models (the Toulmin model being the most prominent of these; Toulmin 1984, 2003), a formal argument consists of several components, including a claim and at least one premise. The claim is the central component and statement, which is justified by premises. Hence, according to Toulmin’s argumentation theory, a good

argument involves a logical structure founded on grounds, claim, and warrant, whereas the grounds are the evidence used to prove a claim (Toulmin 1984). To measure the formal argumentation coherence of arguments as a proxy for objective argumentation quality, we applied the annotation scheme for argumentative knowledge construction (Weinberger and Fischer 2006). This annotation scheme is based on Toulmin’s argumentation theory and has been applied in various studies and is proven to have high objectivity, reliability, and validity (see, e.g., Stegmann et al. 2012). To measure the argumentation’s formal quality, the annotator had to distinguish between (a) unsupported claims, (b) supported claims, (c) limited claims, and (d) supported and limited claims. A more precise description of the scheme can be found in Weinberger and Fischer (2006). Two annotators, who had already participated in the annotation process for our corpus, annotated the participants’ texts based on our annotation guidelines and prior experience. The objective quality of individual users’ argumentation was then defined by the number of arguments that the user wrote. Following Stegmann et al. (2012), only supported, limited, and supported and limited claims were counted as argumentation.

**2. Subjective Quality of Argumentation.** We also capture the subjective quality of argumentation of learners through third-party assessment. The subjective quality of argumentation was annotated by two different annotators. The goal was to subjectively judge how persuasive the given argumentation was on a five-point Likert scale (1: not very persuasive, 5: very persuasive). We took the mean of both annotators as a final variable for the texts’ subjective and objective quality of argumentation.

## Results

In this section, we will present the results of our experimental studies to test our hypothesis. Table 3 presents an overview of the experimental studies with our dynamic argumentation modeling approach.

### H1: Impact of Dynamic Argumentation Modeling vs. Scripted Argumentation Modeling

In study 1, our aim was to evaluate H1 if dynamic argumentation modeling improved persuasive writing performance more so than the scripted argumentation modeling. We used linear regression models (ANOVA) and checked their assumptions visually with a test for normality and a test for homoscedasticity; all assumptions were met. To control for potential effects of interfering variables with our small sample size and to ensure that randomization was successful, we compared the differences in the means of the three constructs included in the presurvey. For all three

**Table 3.** Overview of Experimental Studies with Our Dynamic Argumentation Modeling System

| Study                                                              | <i>n</i> | Objective                                                                                                                                                                                                                                                                                                     | Result                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------------------------------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pilot study: Technical and design evaluations of the artifacts     | 46       | Ensure that the design correctly addresses the needs of the users and that technical artifacts provide high-quality argumentation modeling feedback.                                                                                                                                                          | Moderate agreement for argumentation annotation. Satisfying accuracy and performance of argumentation-mining algorithm. Positive user feedback on the implementation of the learning system. Eye-tracking studies on how students deal with erroneous modeling.                                                                                                                                                  |
| Study 1: Evaluation of H1 in a laboratory experiment               | 54       | Test H1: <i>Whether dynamic argumentation modeling improves persuasive writing performance more so than scripted argumentation modeling.</i>                                                                                                                                                                  | Individuals receiving dynamic argumentation modeling wrote texts with better objective quality of argumentation as well as better subjective quality of argumentation compared with those receiving the benchmark of scripted argumentation modeling.                                                                                                                                                            |
| Study 2: Evaluation of H2 and H3 in a laboratory experiment        | 142      | Test H2: <i>Whether dynamic argumentation modeling improves persuasive writing performance more so than adaptive support; and</i><br>Test H3: <i>whether, compared with adaptive support, dynamic argumentation modeling leads to better persuasive writing performance on both complex and simple tasks.</i> | Dynamic modeling helps learners to write their text with a better objective quality of argumentation as well as better subjective quality of argumentation compared with those using traditionally adaptive support approaches. When facing complex tasks, the positive effect of dynamic modeling compared with adaptive support is even stronger (large effect) than when facing simple tasks (medium effect). |
| Study 3: Evaluation of H4 in a real-life setting over three months | 124      | Test H4: <i>Whether dynamic argumentation modeling on repeated argumentation tasks (over three months) leads to better learning in comparison with static modeling.</i>                                                                                                                                       | Dynamic modeling on learners’ argumentation errors on business model reviews helps them to write texts with a higher quality of argumentation in another argumentation domain (reviews about social debates). Learners receiving dynamic modeling wrote texts with a better objective quality of argumentation than those receiving static or no argumentative modeling.                                         |

constructs, including personal innovativeness, feedback seeking of individuals, and passive argumentative competency, we obtained  $p$ -values greater than 0.05 between the treatment and the control group. The  $p$ -value for personal innovativeness between both groups was  $p = 0.801$ , for feedback-seeking of individuals  $p = 0.624$ , and for passive argumentative competency  $p = 0.375$ , verifying that no significant difference exists between the groups.

The mean number of arguments in texts from participants using *ArgueLearn* was 5.08 (SD = 1.76). For participants using the alternative tool, we counted a mean of 3.2 arguments (SD = 1.51). A linear regression confirmed that the treatment group wrote texts with a statistically significantly higher objective argumentation quality:  $t$ -value =  $-3.622$  and  $p < 0.001$  (see Table 4). For the subjective quality of argumentation, we found that on a five-point Likert scale (1: not very persuasive, 5: very persuasive), texts from the treatment group achieved an average value of 3.38 (SD = 0.96). Participants using the argumentative scripting application wrote texts with a mean subjective quality of argumentation value of 2.79 (SD = 1.19). A linear regression showed that the difference was statistically significant:  $t$ -value =  $-2.654$  and  $p$ -value =  $0.0105$  ( $p < 0.05$ ). This indicates that dynamic modeling helps individuals to

write more convincing texts. The results show that learners receiving dynamic ML-based argumentation modeling wrote texts with better objective quality of argumentation as well as a better subjective quality of argumentation compared with those using the scripted argumentation modeling benchmark approach. For technology acceptance, we calculated each construct’s average. The answers were provided on a five-point Likert scale (1: strongly disagree, 5: strongly agree). *ArgueLearn*’s perceived usefulness was rated with a mean value of 3.48 (SD = 0.58), and its average perceived ease of use was 3.83 (SD = 0.65). The mean value of intention to use as a writing tool was 3.67 (SD = 0.58). The results demonstrate that our participants rated *ArgueLearn* as a dynamic argumentation modeling tool positively compared with midpoints. We included open questions in our survey to discern the participants’ opinions on the tools used and further evaluated our proposed design. The general attitude toward the dynamic modeling was positive; the fast, direct feedback (F6), graph-like visualization of the argumentation structure (F5), and summarizing scores (F3) were highlighted. Minor comments were made on the effect of erroneous predictions and the accuracy of the model (e.g., “I’m not sure how well this algorithm really understands what I’m writing.”).

**Table 4.** Mean and Standard Deviation of Objective and Subjective Argumentation Between Participants Receiving Dynamic Modeling (TG) and Scripted Argumentation Modeling (CG) (Study 1)

| Group                 | <i>n</i> | Objective quality of argumentation <sup>a</sup> |      | Subjective quality of argumentation <sup>a</sup> |      | Mean age | Gender            |
|-----------------------|----------|-------------------------------------------------|------|--------------------------------------------------|------|----------|-------------------|
|                       |          | Mean                                            | SD   | Mean                                             | SD   |          |                   |
| TG: dynamic modeling  | 24       | 5.08                                            | 1.76 | 3.38                                             | 0.96 | 23.8     | 9 female, 15 male |
| CG: scripted modeling | 30       | 3.20                                            | 1.51 | 2.79                                             | 1.19 | 23.03    | 8 female, 22 male |
| Total                 | 54       | <i>t</i> -value = $-3.622$<br>( $p < 0.001$ )   |      | <i>t</i> -value = $-2.654$<br>( $p < 0.05$ )     |      |          |                   |

<sup>a</sup>1 low, 5 high.

## H2 and H3: Impact of Dynamic Argumentation Modeling vs. Adaptive Support and Task Difficulty

In study 2, our aim was to evaluate *H2 whether dynamic argumentation modeling improves persuasive writing performance more so than adaptive support and H3 if, compared with adaptive support, dynamic argumentation modeling leads to better persuasive writing performance on both complex and simple tasks*. Again, we used linear regression models (ANOVA analysis) and checked their assumptions visually. Also, we controlled the differences in the means of the three constructs included in the presurvey. No differences were found between the four groups (e.g., personal innovativeness  $p = 0.61$  or feedback seeking of individuals  $p = 0.805$ ).

For H2 and H3 (see Tables 5–7), we examined the impact of dynamic modeling versus adaptive support while at the same time varying task difficulty (complex versus simple task). The results indicated a clear advantage for dynamic modeling when considering the objective quality of argumentation. For the complex task condition, TG1 (dynamic modeling – complex task) showed a mean of 4.56 (SD = 3.01) in objective quality of argumentation, which was significantly higher than TG3 (adaptive support – complex task), with a mean of 2.87 (SD = 1.80). The Tukey post hoc test confirmed this difference with a large effect size ( $p < 0.001$ ,  $d = 1.1097$ ). This significant disparity highlights the efficacy

of dynamic modeling over adaptive support in improving the objective quality of argumentation, especially in challenging and more complex tasks. For subjective quality of argumentation, we do not observe a significant difference between both groups. Thus, H2 is partially confirmed. Similarly, for the simple task condition, TG2 (dynamic modeling – simple task) had a mean of 3.70 (SD = 3.91) in objective quality of argumentation, surpassing TG4 (adaptive support – simple task), which had a mean of 2.62 (SD = 1.40). The Tukey test revealed in this case a medium-effect size difference between these groups ( $p = 0.0216$ ,  $d = 0.6717$ ), reinforcing the notion that dynamic modeling is more effective than adaptive support regardless of task difficulty. Again, we do not observe a significant difference for both groups considering subjective quality of argumentation. Thus, H3 is also supported partially, although we note the difference in effect sizes for objective quality of argumentation, providing evidence for a more nuanced effect of the dynamic modeling approach for more complex tasks. In summary, the results support H2, demonstrating that dynamic argumentation modeling is more effective than adaptive support in enhancing persuasive writing performance when considering objective quality of argumentation. For H3, the findings suggest that there is a large effect when learners face complex tasks compared with simple tasks (medium effect) when considering objective

**Table 5.** Mean and Standard Deviation of Objective and Subjective Quality of Argumentation (Study 2)

| Group                                | <i>n</i> | Objective quality of argumentation <sup>a</sup> |      | Subjective quality of argumentation <sup>a</sup> |      | Mean age | Gender             |
|--------------------------------------|----------|-------------------------------------------------|------|--------------------------------------------------|------|----------|--------------------|
|                                      |          | Mean                                            | SD   | Mean                                             | SD   |          |                    |
| TG1: dynamic modeling – complex task | 30       | 4.56                                            | 3.01 | 3.46                                             | 1.15 | 24.0     | 11 female, 19 male |
| TG2: dynamic modeling – simple task  | 34       | 3.70                                            | 3.91 | 2.88                                             | 1.25 | 23.34    | 13 female, 20 male |
| TG3: adaptive support – complex task | 41       | 2.87                                            | 1.80 | 2.87                                             | 1.00 | 23.03    | 15 female, 26 male |
| TG4: adaptive modeling – simple task | 37       | 2.62                                            | 1.40 | 2.51                                             | 0.42 | 23.78    | 13 female, 24 male |
| Total                                | 142      |                                                 |      |                                                  |      |          |                    |

<sup>a</sup>1 low, 5 high.

**Table 6.** Mean and Standard Deviation of Subjective Quality of Argumentation Between the Four Groups (Study 2)

| Group                                | N   | Subjective quality of argumentation <sup>a</sup> |      | Tukey HSD test<br>TG2 and Cohen's   | Tukey HSD test<br>TG3 and Cohen's    | Tukey HSD test<br>TG4 and Cohen's  |
|--------------------------------------|-----|--------------------------------------------------|------|-------------------------------------|--------------------------------------|------------------------------------|
|                                      |     | Mean                                             | SD   | d                                   | d                                    | d                                  |
| TG1: dynamic modeling – complex task | 30  | 3.46                                             | 1.15 | p = 0.0822<br>d = 0.5312 (small)    | p = 0.0615<br>d = 0.5689 (medium)    | p = 0.0006***<br>d = 1.101 (large) |
| TG2: dynamic modeling – simple task  | 34  | 2.88                                             | 1.25 | —                                   | p = 0.9999<br>d = 0.004 (negligible) | p = 0.3844<br>d = 0.4066 (small)   |
| TG3: adaptive support – complex task | 41  | 2.87                                             | 1.00 | —                                   | —                                    | p = 0.2530<br>d = 0.426 (small)    |
| TG4: adaptive modeling – simple task | 37  | 2.51                                             | 0.42 | —                                   | —                                    | —                                  |
| Total                                | 142 | p = 0.00163** (p < 0.001)                        |      | ***p < 0.001, **p < 0.01, *p < 0.05 |                                      |                                    |

<sup>a</sup>1 low, 5 high.

quality of argumentation. The interplay between task difficulty and the type of support provided appears to be a crucial factor in enhancing argumentation skills.

**H4: Impact of Dynamic Argumentation Modeling vs. Static Argumentation Modeling**

To test H4, we took the written peer reviews from the posttest of study 2 to measure the students’ objective quality of argumentation. We applied the same annotation scheme for argumentative knowledge construction used in study 1 (Weinberger and Fischer 2006). Two annotators, who had already participated in the annotation process for our corpus, annotated the participants’ texts based on our annotation guidelines and prior experience. As in study 1, we took the mean of the two annotators to judge the texts’ objective and subjective quality of argumentation. We compared the objective quality of argumentation between the written text of the postsurvey (for an overview, see Tables 8 and 9). To clean our data, we deleted all answers of

fewer than three sentences, because an argumentative discourse structure can be reliably measured only from a series of subsequent sentences. Therefore, from our 205 randomly assigned participants, we obtained 124 valid results. We used ANOVA analysis and checked their assumptions visually with a test for normality and a test for homoscedasticity, and all assumptions were met. To control for potential effects of interfering variables and to ensure successful randomization, we compared the differences in the means of the three constructs included in the pretest. For all three constructs, including personal innovativeness, feedback seeking of individuals, and subjective argumentation skills, we received *p*-values larger than 0.05 between the three groups.

The number of formal arguments in texts from participants using the dynamic argumentation modeling tool (TG1) is significantly higher than students receiving static modeling (TG2) and no modeling (CG). The ANOVA analysis confirmed that the TG2 participants

**Table 7.** Mean and Standard Deviation of Objective Quality of Argumentation Between the Four Groups (Study 2)

| Group                                | n   | Objective quality of argumentation <sup>a</sup> |      | Tukey HSD test<br>TG2 and Cohen's   | Tukey HSD test<br>TG3 and Cohen's  | Tukey HSD test<br>TG4 and Cohen's  |
|--------------------------------------|-----|-------------------------------------------------|------|-------------------------------------|------------------------------------|------------------------------------|
|                                      |     | Mean                                            | SD   | d                                   | d                                  | d                                  |
| TG1: dynamic modeling – complex task | 30  | 4.56                                            | 3.01 | p = 0.1308<br>d = 0.4607 (small)    | p < 0.001***<br>d = 1.1097 (large) | p < 0.001***<br>d = 1.3345 (large) |
| TG2: dynamic modeling – simple task  | 34  | 3.70                                            | 3.91 | —                                   | p = 0.1088<br>d = 0.4963 (small)   | p = 0.0216*<br>d = 0.6717 (medium) |
| TG3: adaptive support – complex task | 41  | 2.87                                            | 1.80 | —                                   | —                                  | p = 0.8885<br>d = 0.2014 (small)   |
| TG4: adaptive modeling – simple task | 37  | 2.62                                            | 1.40 | —                                   | —                                  | —                                  |
| Total                                | 142 | p = 0.00002*** (p < 0.001)                      |      | ***p < 0.001, **p < 0.01, *p < 0.05 |                                    |                                    |

<sup>a</sup>1 low, 5 high.

**Table 8.** Mean and Standard Deviation of Objective Quality of Argumentation Between the Three Groups of Our Field Experiment (Study 3)

| Group                 | <i>n</i> | Objective quality of argumentation <sup>a</sup> |      | TukeyHSD test TG2 and Cohen's <i>d</i>     | TukeyHSD test (CG) and Cohen's <i>d</i>      | Mean age | Gender             |
|-----------------------|----------|-------------------------------------------------|------|--------------------------------------------|----------------------------------------------|----------|--------------------|
|                       |          | Mean                                            | SD   |                                            |                                              |          |                    |
| TG1: Dynamic modeling | 41       | 4.04                                            | 1.33 | $p = 0.0150^*$<br>$d = 0.6829$<br>(medium) | $p = 0.001^{**}$<br>$d = 0.7395$<br>(medium) | 25.40    | 21 female, 20 male |
| TG2: Static modeling  | 36       | 3.20                                            | 1.09 | —                                          | $p = 0.8137$<br>$d = 0.1374$<br>(negligible) | 25.70    | 8 female, 28 male  |
| CG: No modeling       | 47       | 3.03                                            | 1.40 | —                                          | —                                            | 25.27    | 18 female, 29 male |
| Total                 | 124      | $p = 0.0009^{***}$ ( $p < 0.001$ )              |      | $^{**}p < 0.01$ , $^*p < 0.05$             |                                              |          |                    |

<sup>a</sup>1 low, 5 high.

wrote texts with a statistically significantly higher objective quality of argumentation:  $p = 0.0009$  ( $p < 0.001$ ). To measure the effects between the groups, we ran a Tukey post hoc test to check for differences in the objective quality of argumentation and confirmed a significant difference between participants using the dynamic ML-based argumentation modeling tool (TG2) and participants receiving static argumentation modeling (TG1) with a  $p$ -value of 0.0150 ( $p < 0.05$ ). Moreover, the test also confirmed that students in TG2 (dynamic modeling with *ArgueLearn*) wrote their text with better objective quality of argumentation than the control group with a  $p$ -value of 0.001 ( $p < 0.01$ ). We also calculated the Cohen's  $d$  between the different groups to measure the effect size. Cohen suggested that  $d = 0.2$  be considered a “small” effect size, whereas 0.5 represents a “medium” effect size and 0.8 a “large” effect size. This means that if two groups' means do not differ by 0.2 standard deviations or more, then the difference is trivial even if statistically significant (Cohen 1988). For the effect between TG2 (dynamic modeling with *ArgueLearn*) and TG1, a  $d = 0.6829$  indicates a medium effect size. The differences between TG2 and CG also indicate a medium effect ( $d = 0.7395$ ) (Cohen 1988) for objective quality of argumentation.

We found no significant difference in subjective quality of argumentation between the groups ( $p = 0.8969$ , see Table 9). Participants using *ArgueLearn* (TG2) wrote

their texts with a subjective quality of argumentation of 3.06 (SD = 0.8077), whereas TG1 students wrote with a subjective quality of argumentation of 2.76 (SD = 0.9140). In the CG texts, we measured a subjective quality of argumentation of 3.07 (SD = 0.9998). These results indicate that dynamic modeling on students' argumentation helps them to write objectively more argumentative texts in another domain (e.g., persuasive reviews about social debates). However, it does not help them to write with higher subjective quality of argumentation levels in other domains. As described above, we also included open questions in our survey to obtain the participants' opinions on *ArgueLearn*'s design and interactivity to further evaluate our functionalities. The general attitude toward *ArgueLearn* was positive. Again, the fast and direct feedback mechanism (artifact 3, F3), the argumentation learning dashboard (artifact 4, F1-F6), and the graph-like visualization of the argumentation structure (artifact 4, F5) were praised several times. However, sometimes *ArgueLearn* did not correctly classify claims and premises, which users suggest needs improving (artifact 3, accuracy of the argumentation mining algorithm).

### Post Hoc Analysis of Learner Expertise

As we controlled in all studies for learner expertise, which is considered as an important aspect for (argumentation) learning outcomes (Marcus et al. 1996, Lin

**Table 9.** Mean and Standard Deviation of Subjective Quality of Argumentation Between the Three Groups of Our Field Experiment (Study 3)

| Group                 | <i>n</i> | Subjective quality of argumentation |      | Tukey HSD test TG2 | Tukey HSD test (CG) |
|-----------------------|----------|-------------------------------------|------|--------------------|---------------------|
|                       |          | Mean                                | SD   |                    |                     |
| TG1: Dynamic modeling | 41       | 3.06                                | 0.80 | $p = 0.8486$       | $p = 0.9886$        |
| TG2: Static modeling  | 36       | 2.76                                | 0.91 | —                  | $p = 0.7842$        |
| CG: No modeling       | 47       | 3.07                                | 0.99 | —                  | —                   |
| Total                 | 124      | $p = 0.8969$                        |      |                    |                     |

et al. 2020), we further explored in a post hoc analysis whether learner expertise has an impact on the effectiveness of our approach. Therefore, we checked all conducted studies for the impact of the variable learner expertise on the argumentation outcomes of students. For this, we assessed the presurvey data from our studies to conduct a mean split for each group of the variable of “argumentation competence” (presurvey, i.e., Flender et al. 1999). We compared the split groups of the conditions (four groups in study 1, eight groups in study 2, and six groups in study 3) with the argumentative outcomes. Across all three studies, we did not find any significant difference between any of the groups on the objective and the subjective argumentation skills (study 1: objective quality of argumentation  $p = 0.40376$ , subjective quality of argumentation  $p = 0.91270$ ; study 2: objective quality of argumentation  $p = 0.1473$ , subjective quality of argumentation  $p = 0.1467$ ; study 3: objective quality of argumentation  $p = 0.6668$ , subjective quality of argumentation  $p = 0.3477$ ).

## Discussion and Implications

### Discussion of Findings

This work makes several contributions to academic research in the fields of argumentation learning, technology-mediated learning, and teaching with IS, especially for the development of skills with artificial intelligence and social cognitive theory. Providing support for individuals with technology-mediated feedback to learn and practice argumentation is a growing and important research field for a skill-based (continuous) education (OECD 2018, WEF 2018). Hence, our objective was to investigate whether dynamic argumentation modeling helps learners increase their argumentation skills across domains compared with the benchmarks of scripted argumentation modeling, adaptive support, static modeling, and no modeling. Based on a socio-technical IS perspective guided by social cognitive theory, we demonstrate how novel learning systems can automate dynamic argumentation modeling for individuals to learn and practice argumentation skills based on their logical argumentation errors irrespective of instructor, time, and location. We empirically examine the role of dynamic argumentation modeling on students’ persuasive writing with two randomized laboratory experiments and a long-term field experiment. The results show that dynamic argumentation modeling helps students to write more persuasive texts compared with scripted modeling (study 1), adaptive support (study 2), and static as well as no modeling (study 3) when considering objective quality of argumentation. Moreover, learners who repeatedly received dynamic argumentation modeling over three months improved their objective argumentation skills significantly in another argumentative writing domain compared with

learners who received static or no argumentation modeling. Furthermore, we provide evidence that the effects of dynamic argumentation modeling are present for both complex tasks and simple tasks (compared with traditional adaptive support approaches when considering objective quality of argumentation). Also, the findings do not show any influence of prior learner expertise on our dependent variables.

### Theoretical Contributions and Practical Implications

Our work directly contributes to the literature on argumentation learning systems (Scheuer et al. 2010), writing support systems (Afrin et al. 2021), and technology-mediated learning with IS (Gupta and Bostrom 2013). In order to illuminate the additional boundary conditions of our results, we underline our findings with qualitative comments from the laboratory experiments (study 1 and study 2) and the field experiment (study 3). Although research has recognized the potential of dynamic modeling for supporting students and educators and possibly transforming institutional activities (such as persuasive essay scoring) (Scheuer et al. 2010, Scheuer 2015, Stab and Gurevych 2017b, Afrin et al. 2021), little attention has been paid to the impact of IS based on ML capabilities (i.e., argumentation mining) and students’ learning outcomes in the field. Our work combines previous findings and theories for the design of technology-mediated learning systems and argumentation mining technology (Bostrom and Heinen 1977, Gupta and Bostrom 2009, Scheuer et al. 2010, Fan et al. 2020, Afrin et al. 2021, Huang et al. 2021) to propose an IT solution for providing dynamic argumentation modeling support to learners. We present an empirical evaluation based on both controlled laboratory studies and a long-term field experiment of a dynamic argumentation modeling system based on ML (Scheuer et al. 2010, Noroozi et al. 2020). Past research has presented general artifact design studies mainly in the fields of NLP and HCI (Scheuer et al. 2012, Scheuer 2015, Lippi and Torroni 2016, Chernodub et al. 2019, Lauscher et al. 2019, Afrin et al. 2021); however, it has lacked rigorous studies about the embedding and impact of these IS on students’ argumentation writing skills (Scheuer 2015, Stab and Gurevych 2017b). In fact, research has shown the impact of scripted argumentation modeling on the objective quality of argumentation, which has been the standard for training argumentation skills at scale (see, e.g., Stegmann et al. 2012, Fischer et al. 2013). We tested the impact of dynamic modeling through our instantiation *ArgueLearn* in two controlled laboratory experiments with 196 learners (study 1 and study 2) and an extensive field experiment with 124 students over three months (study 3). We found that individuals receiving ML-based dynamic argumentation modeling support improved their short-

and long-term persuasive writing skills with better objective quality of argumentation than participants who received scripted argumentation modeling, adaptive support approaches, static modeling, or no modeling support. Our results indicate that learners also write text with a higher subjective quality of argumentation in short-term scenarios. In this context, two learners in the treatment condition of study 1 highlighted the inhibiting effect of dynamic modeling in their learning process:

It was positive that the tool presented which premises support which of my claims and which arguments hung, so to speak, freely in the air.

I really like the visualization of the individual arguments: It is shown which theses are put forward and which arguments are used to support them. This shows what is still lacking in a meaningful argumentation.

However, the subjective argumentation level seems to not significantly improve through our dynamic argumentation modeling system across different learning domains. The effectiveness of dynamic argumentation modeling is not only valuable for students learning argumentation skills and educators judging students' argumentation levels but also for higher education institutions operating in an increasingly complex and competitive environment (Daniel 2015).

Our findings indicate that dynamic argumentation modeling significantly enhances the subjective quality of argumentation even more than adaptive support. This supports the argument presented by Scheuer et al. (2014) and McLaren et al. (2010), who advocated for adaptive support in argumentation learning. However, our study extends this by demonstrating the superiority of dynamic modeling, which is a more sophisticated form of support. Unlike basic adaptive approaches that rely primarily on textual feedback or dashboard-like charts, dynamic modeling, as we implemented, incorporates advanced features like dynamic highlighting and graph-like representations of argumentation structures, focusing on learners' errors.

Theoretically, this aligns with the principles of social cognitive theory, which emphasizes the importance of interactive and dynamic learning environments. As suggested by Guerraoui et al. (2023), adaptive feedback mechanisms like highlighting are insufficient on their own. Our study's results underscore the importance of integrating these mechanisms within a broader, more interactive framework that encourages deeper engagement with the argumentation process. Practically, the incorporation of dynamic elements in argumentation modeling offers a more immersive and responsive learning experience, leading to improved argumentation skills.

The findings for task difficulty suggest that the positive effect when comparing adaptive support and our

approach of dynamic argumentation modeling is even stronger (considering the effect size) when learners face a more complex task. In the context of argumentation learning, complex tasks typically require deeper cognitive engagement, making them ideal for testing the full potential of dynamic argumentation modeling (Sweller et al. 1998, Nussbaum 2008). The enhanced effectiveness of our dynamic modeling approach, especially in the face of challenging tasks, aligns with social cognitive theory, which posits that learning is most effective when it challenges and engages the learner. This theory suggests that dynamic modeling, with its real-time feedback and adaptive learning pathways, is particularly suited to complex tasks where traditional methods, such as static modeling or basic adaptive support, may fall short. Even though, in simpler tasks, we also observe a difference in performance between dynamic modeling and adaptive support, our results indicate that the benefits of sophisticated modeling techniques are more pronounced when students are challenged with complex argumentative structures and ideas.

Our findings show nuanced results on the impact of dynamic argumentation modeling and adaptive support on persuasive writing performance. H2 and H3 were only partially supported because our study showed that dynamic modeling enhances learners' ability to construct arguments with better objective quality in comparison with traditional adaptive support methods. When facing complex tasks, the positive effect of dynamic modeling on the objective quality of argumentation compared with adaptive support is even stronger (large effect) than when facing simple tasks (medium effect). When it comes to the subjective quality of argumentation, however, we did not find significant differences between dynamic and adaptive approaches for simple tasks. This finding suggests that although dynamic modeling effectively supports the objective quality of argumentation, especially in complex scenarios, it may not equally enhance aspects of argumentation that contribute to subjective quality of argumentation across all task difficulties.

One plausible explanation for this phenomenon lies in the nature of the feedback mechanisms provided by the dynamic modeling approach as well as by the adaptive approach. Both feedback mechanisms are based on Toulmin's theory of argumentation (Toulmin 1984), which looks mostly at objective argumentation structures. In the dynamic modeling approach, students experienced support on their objective quality of argumentation, and that's what they improved in both tasks. Although more formal arguments lead to a higher persuasiveness, our results show that there are other factors that seem to influence the persuasiveness of texts. Based on Aristotle's framework of persuasion, also *ethos* (e.g., credibility) and *pathos* (emotions) can lead to convincing argumentation (alongside to *logos* – formal structure).

Specifically, in simpler tasks, the enhancement of objective quality of argumentation through dynamic modeling does not necessarily translate to improvements in ethos and pathos, which are essential for the subjective quality of arguments. This suggests that although enhancing structural and logical components of arguments is vital, a comprehensive approach that also addresses the rhetorical aspects of persuasion is crucial for fostering persuasive writing skills across varying task complexities. Future research could hence, also dive into modeling ethos and pathos in human argumentation to support persuasive writing more holistically.

We also controlled for learner expertise in our analysis as an important boundary condition in self-regulated learning (see, e.g., Zimmerman and Schunk 2001). We observed a consistent pattern; there were no significant differences in both objective and subjective quality of argumentation among learners of varying expertise levels when utilizing our approach. This uniformity in the effectiveness of our approach, regardless of the learner's initial argumentation proficiency, is particularly insightful because prior research suggested that some learners are just gifted and perform better in argumentative reasoning (Means and Voss 1996). It indicates that our approach, grounded in dynamic modeling and social cognitive theory, is capable of catering to a diverse range of learners, effectively supporting their argumentation learning irrespective of their starting skill level. The differential impact of feedback, as proposed by Kalyuga (2007), suggests that learners at varying levels of expertise may benefit differently from dynamic feedback because of their distinct cognitive processing stages. For instance, novices might focus more on understanding the basic structure of arguments, whereas experts might look for more nuanced feedback to refine their argumentative structure provided by dynamic modeling. In contrast, high-expertise learners might find significant advancements harder to achieve, as explained by Kalyuga et al. (2003), suggesting a convergence in learning outcomes despite initial expertise disparities. This effect is typically caused by the expertise-reversal effect that makes feedback practically useless or even detrimental if a learner exhibits high expertise because feedback information adds another layer of (extraneous) information to process (Kalyuga 2007). Nonetheless, the tailored feedback delivered by our dynamic modeling system may contribute to similar learning improvements across expertise levels, thus weakening the expertise-reversal effect. The consistency in outcomes across different expertise levels aligns well with the observed stronger positive effect of our approach in complex tasks. Complex tasks demand a higher level of engagement and cognitive processing, which our system seems to facilitate effectively for all learners. This suggests that the dynamic nature of our system, with its capacity to

provide tailored, real-time feedback and adapt to the learner's needs, is a key factor in its success. Furthermore, argumentation is a less structured task, where the expertise-reversal effect is not as pronounced as in structured tasks (Nievelstein et al. 2013).

Our work contributes to the underlying learning mechanisms rooted in social cognitive theory (Bandura 1977). By enabling positive behavior changes in learning processes through dynamic modeling, our results highlight that providing dynamic models based on ML advances IS research that previously relied mainly on static models (see, e.g., Gupta and Bostrom 2013 and Sullivan et al. 2022). This result is especially novel for argumentation skill learning and skills in general because past work has mainly developed theories of learning by examples and by doing (e.g., ACT-R theory and SimStudent theory) in STEM context. This project demonstrates that this theory generalizes outside of math and science domains to argumentative writing (see, e.g., Anderson et al. 1995 and Matsuda et al. 2015). Past research on dynamic argumentation learning has focused mostly on general design elements for adaptive argumentation monitoring, for example, based on in-text highlighting (Scheuer et al. 2010, Chernodub et al. 2019, Wang et al. 2020, Afrin et al. 2021), whereas perspectives on social cognitive learning theory with an interdisciplinary IS research perspective were absent. Moreover, past argumentation learning theories, such as representational guidance theory (Suthers and Hundhausen 2001, Suthers 2003), focused exclusively on argumentation representation forms (e.g., supporting argumentation learning by providing representations of argumentation structures with the objective of stimulating and improving individual reasoning, collaboration, and learning) (Suthers and Hundhausen 2001; Scheuer et al. 2010, 2012). Our research provides novel insights into how learners' argumentative writing based on dynamic modeling may be supported in combination with a socio-technical design embedding. This provides a leap forward to extend social cognitive theory by utilizing ML-based feedback systems. Our study provides novel experimental findings for the impact of these systems on skills. This is also exemplified in the following two comments by learners from study 3:

It was interesting to see the system acknowledges claims and premises.

This is the first time I have used such a tool. Through the analysis, I have been partially made aware of gaps in my argumentation style.

Educational theories in general (such as social cognitive theory) and argumentation learning theories specifically (Scheuer et al. 2010, Bjork et al. 2013, Metcalfe 2017, Lawrence and Reed 2019) demonstrate the potential to investigate dynamic modeling as a new genre of

educational IS to provide natural observational learning processes for the maximum impact of IS in educational settings, thus advancing the rather static view of IS research on behavioral modeling (Yi and Davis 2003, Sullivan et al. 2022). Learners stand to benefit from the application of ML-based feedback as well as educators and educational institutions, which could increase their efficiency and effectiveness. For example, dynamic argumentation modeling could also be utilized to automate argumentative essay correction on a deeper granular level (subsentence-based). Educators may rely on ML-based approaches to receive reliable support in scoring student-written essays' argumentation. The theory-based view of our proposed IT solution as a system class for technology-mediated learning can encourage educational IS researchers to further investigate, implement, and drive the research and functionalities of ML-based argumentation learning systems. The systematic conceptualization in a service-oriented architecture provides a solid foundation for further development and implementation of such systems (Elshan et al. 2023). In particular, the module-based composition can encourage research in different domains and disciplines to contribute to an (argumentation) skill-based education in the future (e.g., for computer linguistics by providing more student-written corpora and models or for HCI by investigating interaction designs for novel skill training). Moreover, we exemplify how to embed argumentation mining technology in an IS context. Argumentation mining is a powerful method and toolset for investigating and contributing to various IS research phenomena, for example, in the context of opinion mining, decision support systems, logical online discussions, or social media analysis (Abbasi et al. 2008, 2018; Deng et al. 2019; Lawrence and Reed 2019). However, argumentation mining has not been widely adopted in IS research (Skiera et al. 2022). The resulting knowledge is valid not only for our specific case but also for further use cases in adaptive argumentation learning. For instance, the architecture and the modules may be applied in courses dealing with content other than business models or other languages; the back-end algorithm must simply be adapted to the other scenario. As described, multiple corpora and argumentation mining models can be easily embedded as a module in the service-oriented architecture for dynamic argumentation feedback systems (e.g., for English students' essays (Stab and Gurevych 2017b) or English law cases (Mochales Palau and Ieven 2009). The user interface and the overall system design need not be adapted for these use cases. It is also possible to transfer the design knowledge to pedagogical scenarios that target other skills. For instance, for general feedback skills or empathy skills, a similar learning system can be used. However, in this case, the system design may require partial revision (e.g., the graph visualization or dashboard may need to be adapted).

## Limitations, Future Research, and Conclusion

Our research has several limitations and boundary conditions that bear potential for possible future research. First, our approach was evaluated on learners in the context of higher education. Although it is reasonable to assume that transferability to other cases is possible without major changes, we cannot prove it with our research design. Second, our studies were limited in sample size. Further empirical evaluations are necessary to replicate the results in other educational domains and additional sample sizes to further evaluate the effects of dynamic modeling on students' skills. Third, we aim to highlight our study's ethical limitations; regarding the implementation of our intelligent learning system, we have no desire to replace human educators, because we believe that skilled teachers will—for the foreseeable future—be able to provide better dynamic modeling than ML. Concerns about receiving feedback from artificial intelligence were also evident in some learner comments, for example, like this exemplary comment from a student in study 3:

I am not sure if this tool is really useful. In my opinion, argumentative texts cannot be evaluated by AI.

However, we hope that through our system, human educators can focus more on detailed questions and devote more time to difficult cases. For example, following the logic of hybrid intelligence, educators or peers could focus more on the quality of content of the argumentation, whereas an ML-based learning system targets the structural elements of the argumentation. This potential benefit of AI-based systems for learning processes is especially evident given the uprise of large language models such as OpenAI's GPT4 or the ChatGPT system that make it necessary to think about the beneficial usage of ML-based learning systems instead of forbidding them (Mollick and Mollick 2022). Fourth, we also perceive several data privacy concerns regarding the integration of our tool into a common writing editor because personal data (e.g., about students' argumentation skills) might be exposed to third parties or systematically collected. Hence, we recommend future discussions on dealing with the trade-off of making novel user-centered learning applications widely accessible and easy to use without exposing learner data (Dickhaut et al. 2023). Nevertheless, because our system is based entirely on native libraries and frameworks, we can ensure that our architecture ensures that no personal data leave the university server's infrastructure.

Fifth, and most importantly, a natural limitation arises from the nature of dynamic ML-based modeling and possibly biased or erroneous advice. Although we

took several measures to counteract and control for possible harm of our dynamic ML-based modeling tool, we cannot explicitly expel any negative effects. This can be seen, for example, in this exemplary comment stated by a learner (study 1):

I have the feeling that depending on the sentence structure, the tool does not recognize whether an argument is also supported with facts or examples.

To counter possible bias and erroneous predictions, we took several measures. First of all, we focused on the possibility of erroneous predictions of dynamic ML-based modeling from the very beginning of the core design of our study. For example, we explicitly asked in our 30 user interviews about the options of erroneous feedback and incorporated that into the user interaction of our tool design. Moreover, we controlled for the effects of erroneous modeling on students in 13 eye-tracking studies and then qualitative comments of all experiments (study 1, study 2, and study 3). In fact, based on the user requirements derived from our 30 interviews and the known possibility of erroneous recommendations, we decided, for example, to not explicitly recommend students to change their argumentation if they had an error (e.g., “please add a premise here”) but instead just indicated their argumentative error. Also, the way the system is introduced to students, as well as how the explanations and help texts are written, set the mindset for the student that *ArgueLearn* is a “learning” prototype (“also a student like you, that is learning and possibly sometimes making unintended mistakes”) to provide the right framing. Nevertheless, we controlled for any harm and problems in the qualitative comments of students in the survey after the treatment of both studies. Although we did not find a significant number of negative comments (most students did not mention the effect of erroneous predictions on their learning task), some students mentioned erroneous predictions (as exemplified by the user comment above).

We believe that as learners become more proficient with novel technologies (e.g., ChatGPT; Knoth et al. 2024), they may be able to identify erroneous feedback and might develop strategies to mitigate the negative impact of such errors on their learning. Interestingly, this tendency can also be regarded in the literature of algorithmic appreciation and aversion in the context of erroneous AI advice (Dietvorst et al. 2018, Fuegener et al. 2021, Schmitt et al. 2021). Nonetheless, we expect that learning to deal with imperfect AI advice will be the future when interacting with these systems. In consequence, achieving AI literacy will be the key for individuals, and future research should take a closer look at how we can improve AI literacy to improve the outcomes of AI-based systems (Tolzin et al. 2024).

Additionally, we performed an ex post Word Embedding Association Test (WEAT) analysis (Caliskan et al. 2017) and the German adaptation of WEAT (Kurpicz-Briki 2020) as a commonly used methodology to assess conceptual, racial, and gender bias in different parts of the NLP pipeline of our corpus to control for systematic biases (Hovy and Prabhumoye 2021). We found that our collected corpus does not reveal many biases in using WEAT co-occurrence analysis or GloVe models.<sup>5</sup>

Finally, we aim to call for future research on the effects of dynamic argumentation modeling on different groups of users. For example, Noroozi et al. (2023) found that female students provided better justifications for problems identified in peer review, more constructive reviews, and higher-quality peer reviews overall than males. Although we did not find gender differences for any variable in our experiment, it may be interesting to examine the effect of dynamic modeling on various demographic variables to control for bias and fairness in our dynamic ML-based modeling. Moreover, although dynamic modeling may have a positive effect on the majority of students, it may also be demotivating for some minority groups, for example, low-performing students, although we did not see any performance effects of learner expertise. Even though we did not see any negative effects in our sample size, future research should shed additional light on the unintended consequences of dynamic argumentation modeling and should investigate whether scaffolding procedures (Janson et al. 2020) are necessary for different user groups with different proficiency levels.

In conclusion, our research offers empirical and design knowledge to further improve educational feedback applications based on intelligent algorithms. With further advances in NLP and ML, we hope our work will encourage researchers to design more intelligent feedback systems for other learning scenarios or skills and thus contribute to the OECD Learning Framework 2030 toward a skill-based education related to transformative competencies such as reconciling tensions and dilemmas.

## Acknowledgments

The authors thank all earlier collaborators on this project such as Christina Niklaus, Matthias Cetto, Niels Pinkward, Antonia Tolzin, Roman Rietsche, and Siegfried Handschuh. Also, the authors acknowledge all students who took part in this study. Lastly, the authors thank the senior editor Gautam Pant and the associate editor Hong Gu as well as the anonymous reviewers for their constructive feedback, trust, and guidance in the revision process.

## Endnotes

<sup>1</sup> More information about the inter-annotator agreements of our argumentation corpus of 1,000 student-written texts, the trained ML model, and its tuned accuracy can be found in Wambsganss et al. (2020b).

<sup>2</sup> The interview guideline consists of 29 questions focusing on the learner-interaction and the argumentation learning process of

students. Each interview lasted around 30 to 50 minutes. The interviewed students were between 22 and 28 years old and were all students of economics, computer science, or psychology; 13 were male, and 17 were female.

<sup>3</sup> The annotation guidelines as well as the entire corpus can be freely accessed at [https://github.com/thiemowa/argumentative\\_student\\_peer\\_reviews](https://github.com/thiemowa/argumentative_student_peer_reviews).

<sup>4</sup> The experimental studies are in line with the ethical guidelines of our university. Moreover, the students gave their consent to utilize the anonymized data for scientific purposes.

<sup>5</sup> The detailed bias analysis on our corpus as well as the methodology are published in Wambsganss et al. (2022a).

## References

- Abbasi A, Chen H, Salem A (2008) Sentiment analysis in multiple languages. *ACM Trans. Inform. Systems* 26(3):1–34.
- Abbasi A, Zhou Y, Deng S, Zhang P (2018) Text analytics to support sense-making in social media: A language-action perspective. *MIS Quart.* 42(2):427–464.
- Afrin T, Kashefi O, Olshefski C, Litman D, Hwa R, Godley A (2021) Effective interfaces for student-driven revision sessions for argumentative writing. *Proc. 2021 CHI Conf. Human Factors Comput. Systems* (ACM, New York), 1–13.
- Agarwal R, Karahanna E (2000) Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quart.* 24(4):665.
- Alavi M, Leidner DE (2001) Research commentary: Technology-mediated learning - A call for greater depth and breadth of research. *Inform. Systems Res.* 12(1):1–10.
- Anderson K (1986) Computer assisted instruction: An overview. *Proc. Hawaii Internat. Conf. System Sci.* 3(October):4–10.
- Anderson JR, Corbett AT, Koedinger KR, Ray P (1995) Cognitive tutors: Lessons learned. *J. Learn. Sci.* 4(2):167–207.
- Ashford SJ (1986) Feedback-seeking in individual adaptation: A resource perspective. *Acad. Management J.* 29(3):465–487.
- Bandura A (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* 84(2):191–215.
- Bandura A (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice-Hall, Englewood Cliffs, NJ).
- Bandura A (1991) Social cognitive theory of self-regulation. *Organ. Behav. Human Decision Processes* 50(2):248–287.
- Bandura A (2001) Social cognitive theory: An agentic perspective. *Annu. Rev. Psychol.* 52:1–26.
- Bauman K, Tuzhilin A (2018) Recommending remedial learning materials to students by filling their knowledge gaps. *MIS Quart.* 42(1):313–332.
- Bell BS, Kozlowski SWJ (2008) Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *J. Appl. Psychol.* 93(2):296–316.
- Bjork RA, Dunlosky J, Kornell N (2013) Self-regulated learning: Beliefs, techniques, and illusions. *Annu. Rev. Psychol.* 64(November 2012):417–444.
- Bostrom RP, Heinen JS (1977) MIS problems and failures: A socio-technical perspective. Part I: The causes. *MIS Quart.* 1(3):17.
- Brinton CG, Rill R, Ha S, Chiang M, Smith R, Ju W (2015) Individualization for education at Scale: MIIC design and preliminary evaluation. *IEEE Trans. Learn. Technol.* 8(1):136–148.
- Brown KG, Howardson G, Fisher SL (2016) Learner control and e-learning: Taking stock and moving forward. *Annu. Rev. Organ. Psychol. Organ. Behav.* 3:267–291.
- Byrne RMJ (1989) Everyday reasoning with conditional sequences. *Quart. J. Exp. Psychol. Section A* 41(1):141–166.
- Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186.
- Carter MJ, Ste-Marie DM (2017) An interpolated activity during the knowledge-of-results delay interval eliminates the learning advantages of self-controlled feedback schedules. *Psychol. Res.* 81(2):399–406.
- Chernodub A, Oliynyk O, Heidenreich P, Bondarenko A, Hagen M, Biemann C, Panchenko A (2019) TARGER: Neural argument mining at your fingertips. *Proc. 57th Annual Meeting Assoc. Comput. Linguistics: System Demonstrations* (Association for Computational Linguistics, Kerrville, TX), 195–200.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Routledge, New York).
- Compeau DR, Higgins CA (1995) Application of social cognitive theory to training for computer skills. *Inf. Syst. Res.* 6(2):118–143.
- Daniel B (2015) Big data and analytics in higher education: Opportunities and challenges. *Br. J. Ed. Technol.* 46(5):904–920.
- De Groot R, Drachman R, Hever R, Schwartz B, Hoppe U, Harrer A, De Laat M, Wegerif R, McLaren BM, Baurens B (2007) Computer supported moderation of E-discussions: The ARGU-NAUT approach. *Proc. Conf. Comput.-Supported Collaborative Learn.* (CSCL 2007) (ACM, New York).
- Deng S, Zhou Y, Zhang P, Abbasi A (2019) Using discussion logic in analyzing online group discussions: A text mining approach. *Inform. Management* 56(4):536–551.
- Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conf. North American Chapter Assoc. Comput. Linguistics: Human Language Technologies*, vol. 1 (Long and Short Papers) (Association for Computational Linguistics, Kerrville, TX).
- Diana N (2018) Leveraging educational technology to improve the quality of civil discourse. *Internat. Conf. Artificial Intelligence Ed.* (Springer Verlag, Cham, Switzerland), 517–520.
- Dickhaut E, Janson A, Söllner M, Leimeister JM (2023) Lawfulness by design – Development and evaluation of lawful design patterns to consider legal requirements. *Eur. J. Inf. Syst.*, ePub ahead of print March 1, <https://doi.org/10.1080/0960085X.2023.2174050>.
- Dietvorst B, Simmons J, Massey C (2018) Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Sci.* 64:1155–1170.
- Dillenbourg P (2002) Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three Worlds of CSCL. Can We Support CSCL?* (Open Universiteit Nederland, Heerlen, Netherlands), 61–91.
- Dillenbourg P, Hong F (2008) The mechanics of CSCL macro scripts. *Int. J. Computer Supported Collab. Learn.* 3(1):5–23.
- Dillenbourg P, Järvelä S, Fischer F (2009) The evolution of research on computer-supported collaborative learning. Balacheff N, Ludvigsen S, de Jong T, Lazonder A, Barnes S, eds. *Technology-Enhanced Learning: Principles and Products* (Springer Netherlands, Dordrecht, Netherlands), 3–19.
- Duschl RA, Osborne J (2002) Supporting and promoting argumentation discourse in science education. *Stud. Sci. Ed.* 38(1):39–72.
- Elshan E, Ebel P, Söllner M, Leimeister JM (2023) Leveraging low code development of smart personal assistants: An integrated design approach with the SPADE method. *J. Management Inform. Syst.* 40(1):96–129.
- Eom SB, Wen HJ, Ashill N (2006) The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sci. J. Innovative Ed.* 4(2):215–235.
- Fadel C, Bialik M, Trilling B (2015) *Four-Dimensional Education: The Competencies Learners Need to Succeed* (Center for Curriculum Redesign, Boston).
- Fan YC, Wang TH, Wang KH (2020) Studying the effectiveness of an online argumentation model for improving undergraduate students' argumentation ability. *J. Comput. Assisted Learn.* 36(4):526–539.
- Fischer F, Kollar I, Stegmann K, Wecker C (2013) Toward a script theory of guidance in computer-supported collaborative learning. *Ed. Psychol.* 48(1):56–66.

- Flender J, Christmann U, Groeben N (1999) Entwicklung und erste Validierung einer Skala zur Erfassung der passiven argumentativ-rhetorischen Kompetenz. *Z. Differ. Diagn. Psychol.* 20(4):309–325.
- Flesch R (1943) Marks of readable style; a study in adult education. *Teachers College Contributions to Education* 897.
- Florou E, Konstantopoulos S, Kukurikos A, Karampiperis P (2013) Argument extraction for supporting public policy formulation. *Proc. 7th Workshop Language Tech. Cultural Heritage Soc. Sci. Humanities* (Association for Computational Linguistics, Stroudsburg, PA), 49–54.
- Fuegener A, Grahl J, Gupta A, Ketter W (2021) Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Quart.* 45(3):1527–1556.
- Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. *MIS Quart.* 37(2):337–356.
- Guerraoui C, Reiser P, Inoue N, Mim FS, Naito S, Choi J, Inui K (2023) Teach me how to improve my argumentation skills: A survey on feedback in argumentation. Preprint, submitted July 28, <https://arxiv.org/abs/2307.15341>.
- Gupta S, Bostrom R (2013) An investigation of the appropriation of technology-mediated training methods incorporating enactive and collaborative learning. *Inform. Systems Res.* 24(2): 454–469.
- Gupta S, Bostrom RP (2009) Technology-mediated learning: A comprehensive theoretical model. *J. Assoc. Inform. Systems* 10(9):686–714.
- Guo K, Zhong Y, Li D, Chu SKW (2023) Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Comput. Ed.* 203(15):104862.
- Han J, Kim KH, Rhee W, Cho YH (2021) Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Comput. Ed.* 163:104041.
- Hattie J, Timperley H (2007) The power of feedback. *Rev. Ed. Res.* 77(1):81–112.
- Head A, Lo K, Kang D, Fok R, Skjonsberg S, Weld DS, Hearst MA (2021) Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols. *Proc. 2021 CHI Conf. Human Factors Comput. Systems (CHI '21)* (Association for Computing Machinery, New York), 1–18.
- Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. *Design Sci. IS Res. MIS Quart.* 28(1):75.
- Hone KS, El Said GR (2016) Exploring the factors affecting MOOC retention: A survey study. *Comput. Ed.* 98:157–168.
- Hovy D, Prabhumoye S (2021) Five sources of bias in natural language processing. *Lang. Linguist. Compass.* 15(8):e12432.
- Huang CJ, Chang SC, Chen HM, Tseng JH, Chien SY (2016) A group intelligence-based asynchronous argumentation learning-assistance platform. *Interact. Learn. Environ.* 24(7):1408–1427.
- Huang N, Zhang J, Burtch G, Li X, Chen P (2021) Combating procrastination on massive online open courses via optimal calls to action. *Inform. Systems Res.* 32(2):301–317.
- Janson A, Söllner M, Leimeister JM (2020) Ladders for learning: Is scaffolding the key to teaching problem-solving in technology-mediated learning contexts? *Acad. Management Learn. Ed.* 19(4):439–468.
- Jonassen DH, Kim B (2010) Arguing to learn and learning to argue: Design justifications and guidelines. *Ed. Technol. Res. Dev.* 58(4): 439–457.
- Jung R, Lehrer C (2017) Guidelines for education in business and information systems engineering at tertiary institutions. *Bus. Inform. Systems Engrg.* 59(3):189–203.
- Kalyuga S (2007) Expertise reversal effect and its implications for learner-tailored instruction. *Ed. Psychol. Rev.* 19(4):509–539.
- Kalyuga S, Ayres P, Chandler P, Sweller J (2003) The expertise reversal effect. *Ed. Psychol.* 38(1):23–31.
- Kim D, Benbasat I (2006) The effects of trust-assuring arguments on consumer trust in Internet stores: Application of Toulmin's model of argumentation. *Inform. Systems Res.* 17(3):286–300.
- Knoth N, Tolzin A, Janson A, Leimeister JM (2024) AI literacy and its implications for prompt engineering strategies. *Comput. Ed. Artificial Intelligence* 9(2):100225.
- Koschmann T, ed. (1996) *CSCL: Theory and Practice of an Emerging Paradigm* (Lawrence Erlbaum Associates, Inc., Mahwah, NJ).
- Kuhn D (1992) Thinking as argument. *Harv. Ed. Rev.* 62(2):155–179.
- Kuhn D (1993) Science as argument: Implications for teaching and learning scientific thinking. *Sci. Ed.* 77(3):319–337.
- Kulik JA, Fletcher JD (2016) Effectiveness of intelligent tutoring systems: A meta-analytic review. *Rev. Ed. Res.* 86(1):42–78.
- Kurpicz-Briki M (2020) Cultural differences in bias? Origin and gender bias in pre-trained German and French word embeddings. *Proc. 5th Swiss Text Analytics Conf. (SwissText) & 16th Conf. Natl. Language Processing (KONVENS)* (Zurich, Switzerland).
- Lauscher A, Glavaš G, Eckert K (2019) ArguminSci: A tool for analyzing argumentation and rhetorical aspects in scientific writing. *Proc. 5th Workshop Argument Mining* (Association for Computational Linguistics, Stroudsburg, PA), 22–28.
- Lawrence J, Reed C (2019) Argument mining: A survey. *Comput. Linguist.* 45(4):765–818.
- Lederman D (2018) New data: Online enrollments grow, and share of overall enrollment grows faster. Retrieved November 12, <https://www.insidehighered.com/digital-learning/article/2018/11/07/new-data-online-enrollments-grow-and-share-overall-enrollment>.
- Lehmann T, Hähnlein I, Ifenthaler D (2014) Cognitive, metacognitive and motivational perspectives on prelection in self-regulated online learning. *Comput. Human Behav.* 32:313–323.
- Leidner DE, Jarvenpää SL (1995) The use of information technology to enhance management school education: A theoretical view. *MIS Quart.* 19(3):265–291.
- Leung ACM, Santhanam R, Kwok RCW, Yue WT (2022) Could gamification designs enhance online learning through personalization? Lessons from a field experiment. *Inform. Systems Res.* 34(1): 27–49.
- Lin YR, Fan B, Xie K (2020) The influence of a web-based learning environment on low achievers' science argumentation. *Comput. Ed.* 151:103860.
- Lin YR, She HC, Yang WT, Huang KY (2015) Online collaborative learning for improving argumentation of student with different levels of science prior knowledge. *Proc. IEEE 15th Internat. Conf. Adv. Learn. Tech. Adv. Tech. Supporting Open Access Formal Informal Learn., ICALT 2015* (IEEE, Piscataway, NJ), 214–215.
- Lippi M, Torroni P (2016) MARGOT: A web server for argumentation mining. *Expert Syst. Appl.* 65:292–303.
- Marcus SL, Rips LJ (1979) Conditional reasoning. *J. Verbal Learn. Verbal Behav.* 18(2):199–223.
- Marcus N, Cooper M, Sweller J (1996) Understanding instructions. *J. Ed. Psychol.* 88(1):49–63.
- Matsuda N, Cohen WW, Koedinger KR (2015) Teaching the teacher: Tutoring SimStudent leads to more effective cognitive tutor authoring. *Int. J. Artificial Intelligence Ed.* 25(1):1–34.
- Mayer RE (2014) Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, 2nd ed. (Cambridge University Press, Cambridge, UK), 43–71.
- McAlister S, Ravenscroft A, Scanlon E (2004) Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC. *J. Comput. Assisted Learn.* 20:194–204.
- McLaren BM, Scheuer O, Mikšátko J (2010) Supporting collaborative learning and e-discussions using artificial intelligence techniques. *Int. J. Artificial Intelligence Ed.* 20(1):1–46.
- Metcalfe J (2017) Learning from errors. *Annu. Rev. Psychol.* 68:465–489.
- Means ML, Voss J F (1996) Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cogn. Instr.* 14:139–179.
- Mochales R, Moens MF (2011) Argumentation mining. *Artificial Intelligence Law.* 19(1):1–22.

- Mochales Palau R, Ieven A (2009) Creating an argumentation corpus: Do theories apply to real arguments? A case study on the legal argumentation of the ECHR. *Proc. Twelfth Internat. Conf. Artificial Intelligence Law (ICAIL 2009)* (ACM, New York), 21–30.
- Mollick ER, Mollick L (2022) New modes of learning enabled by AI chatbots: Three methods and assignments. Preprint, submitted December 23, <http://dx.doi.org/10.2139/ssrn.4300783>.
- Nievelstein F, van Gog T, van Dijk G, Boshuizen HPA (2013) The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemp. Ed. Psychol.* 38(2):118–125.
- Niknejad N, Ismail W, Ghani I, Nazari B, Bahari M, Hussin ARBC (2020) Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation. *Inform. Systems* 91:101491.
- Noetel M, Griffith S, Delaney O, Harris NR, Sanders T, Parker P, Cruz B del P, Lonsdale C (2022) Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Rev. Educational Res.* 92(3):413–454.
- Noroozi O, Dehghanzadeh H, Talaei E (2020) A systematic review on the impacts of game-based learning on argumentation skills. *Entertain. Comput.* 35:100369.
- Noroozi O, Kirschner PA, Biemans HJA, Mulder M (2018) Promoting Argumentation competence: Extending from first- to second-order scaffolding through adaptive fading. *Ed. Psych. Rev.* 30(1): 153–176.
- Noroozi O, Banihashem SK, Taghizadeh Kerman N, Parvaneh Akhteh Khaneh M, Babayi M, Ashrafi H, Biemans HJA (2023) Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interact. Learn. Environ.* 31(10):6302–6316.
- North Whitehead A, Russell B (1910) *Principia Mathematica* (Cambridge University Press, Cambridge, UK).
- Nussbaum EM (2008) Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *J. Ed. Psychol.* 100(3), 549–565.
- Nussbaum EM, Winsor DL, Aquilino Y, Poliquin AM (2007) Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. *Int. J. Comput. Supported Collab. Learn.* 2(4):479–500.
- Nye BD (2014) Barriers to ITS adoption: A systematic mapping study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Springer Verlag, Cham, Switzerland), 583–590.
- OECD (2018) The future of education and skills - Education 2030. OECD, Paris.
- Osborne JF, Henderson JB, MacPherson A, Szu E, Wild A, Yao SY (2016) The development and validation of a learning progression for argumentation in science. *J. Res. Sci. Teach.* 53(6):821–846.
- Palau RM, Moens MF (2009) Argumentation mining: The detection, classification and structure of arguments in text. *Proc. 12th Internat. Conf. Artificial Intelligence Law* (Association for Computing Machinery, New York), 98–107.
- Pinkwart N, Ashley K, Lynch C, Alevan V (2009) Evaluating an intelligent tutoring system for making legal arguments with hypotheticals. *Internat. J. Artificial Intelligence Ed.* 19(4):401–424.
- Ravenscroft A, McAlister S, Sagar M (2010) Digital dialogue games and InterLoc: A deep learning design for collaborative argumentation on the web. *Educational Technologies for Teaching Argumentation Skills* (Bentham Science Publishers, Sharjah, U.A.E.), 277–315.
- Roediger HL, Karpicke JD (2006) Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol. Sci.* 17(3): 249–255.
- Rosé C, Wang YC, Cui Y, Arguello J, Stegmann K, Weinberger A, Fischer F (2008) Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *Int. J. Comput. Supported Collab. Learn.* 3(3):237–271.
- Santhanam R, Liu D, Shen WCM (2016) Gamification of technology-mediated training: Not all competitions are the same. *Inform. Systems Res.* 27(2):453–465.
- Santhanam R, Sasidharan S, Webster J (2008) Using self-regulatory learning to enhance E-learning-based information technology training. *Inform. Systems Res.* 19(1):26–47.
- Scheuer O (2015) Toward adaptive argumentation learning systems. Thesis, Universität des Saarlandes, Saarbrücken, Germany.
- Scheuer O, Loll F, Pinkwart N, McLaren BM (2010) Computer-supported argumentation: A review of the state of the art. *Int. J. Comput. Supported Collab. Learn.* 5(1):43–102.
- Scheuer O, McLaren BM, Loll F, Pinkwart N (2012) Automated analysis and feedback techniques to support and teach argumentation: A survey. *Ed. Tech. Teaching Argumentation Skills* 2:71–124.
- Scheuer O, McLaren BM, Weinberger A, Niebuhr S (2014) Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instr. Sci.* 42(2):127–157.
- Schmitt A, Wambsganss T, Janson A, Söllner M (2021) Toward a trust reliance paradox? Exploring the gap between perceived trust in and reliance on algorithmic advice. *Forty Second Internat. Conf. Inform. Systems (Austin, TX)*, 1–17.
- Schunk DH (2012) *Learning Theories: An Educational Perspective*, 6th ed. (Pearson, Boston).
- Seaman JE, Allen IE, Seaman J (2018) Grade increase: Tracking distance education in the United States. Report, Higher Education Reports - Babson Survey Research Group, Babson Park, MA.
- Shehab HM, Nussbaum EM (2015) Cognitive load of critical thinking strategies. *Learn. Instr.* 35(3):51–61.
- Sidorova A, Evangelopoulos N, Valacich JS, Ramakrishnan T (2008) Uncovering the intellectual core of the information systems discipline. *MIS Quart.* 32(3):467–482.
- Skiera B, Yan S, Daxenberger J, Dombois M, Gurevych I (2022) Using information-seeking argument mining to improve service. *J. Serv. Res.* 25(4):537–548.
- Slonim N, Bilu Y, Alzate C, Bar-Haim R, Bogin B, Bonin F, Choshen L, et al. (2021) An autonomous debating system. *Nature* 591(7850): 379–384.
- Smolensky P, Fox B, King R, Lewis C (1988) Computer-aided reasoned discourse or, how to argue with a computer. Guindon R, ed. *Cognitive Engineering in the Design of Human-Computer Interaction and Expert Systems* (Taylor & Francis Group, Amsterdam), 109–162.
- Soloway E, Guzdial M, Hay KE (1994) Learner-centered design: The challenge for HCI in the 21st century. *Interactions* 1(2):36–48.
- Song Y, Heilman M, Klebanov BB, Deane P (2014) Applying argumentation schemes for essay scoring. *Proc. First Workshop Argumentation Mining* (Association for Computational Linguistics, Stroudsburg, PA).
- Stab C, Gurevych I (2014) Annotating argument components and relations in persuasive essays. *Proc. COLING 2014 25th Internat. Conf. Comput. Linguistics Tech. Papers* (Dublin City University and Association for Computational Linguistics, Stroudsburg, PA), 1501–1510.
- Stab C, Gurevych I (2017a) Recognizing insufficiently supported arguments in argumentative essays. Lapata M, Blunsom P, Koller A, eds. *Proc. 15th Conf. Eur. Chapter Assoc. Computat. Linguistics: Volume 1, Long Papers* (Association for Computational Linguistics, Stroudsburg, PA), 980–990.
- Stab C, Gurevych I (2017b) Parsing argumentation structures in persuasive essays. *Comput. Linguist.* 43(3):619–659.
- Stegmann K, Wecker C, Weinberger A, Fischer F (2012) Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instr. Sci.* 40(2):297–323.
- Sullivan Y, Davis F, Koh C (2022) Executive functions and information systems learning. *MIS Quart.* 46(2):813–880.

- Suthers DD (2003) Representational guidance for collaborative inquiry. *Arguing to Learn* (Springer, Dordrecht, Netherlands), 27–46.
- Suthers DD, Hundhausen CD (2001) *European Perspectives on Computer-Supported Collaborative Learning* (University of Maastricht, Maastricht, Netherlands).
- Suthers DD, Hundhausen CD (2003) An experimental study of the effects of representational guidance on collaborative learning processes. *J. Learn. Sci.* 12(2):183–218.
- Sweller J, van Merriënboer JJG, Paas FGWC (1998) Cognitive architecture and instructional design. *Ed. Psychol. Rev.* 10(3):251–296.
- Tolzin A, Knott N, Janson A (2024) Worked examples to facilitate the development of prompt engineering skills. *32nd Eur. Conf. Inform. Systems (ECIS)* (AIS Electronic Library, Atlanta).
- Topi H (2018) Using competencies for specifying outcome expectations for degree programs in computing: Lessons learned from other disciplines. *2018 SIGED Internat. Conf. Inform. Systems Ed. Res.* (AIS Electronic Library, Atlanta).
- Toulmin SE (1984) *Introduction to Reasoning* (Macmillan, New York).
- Toulmin SE (2003) *The Uses of Argument: Updated Edition* (Cambridge University Press, Cambridge, UK).
- Tse C, Roof K (2021) Coursera prices IPO at top of range to raise \$519 Million - Bloomberg. Retrieved April 27, <https://www.bloomberg.com/news/articles/2021-03-31/online-educator-coursera-s-top-of-range-ipo-raises-519-million>.
- Tuunanen T, Peffers K (2018) Population targeted requirements acquisition. *Eur. J. Inf. Syst.* 27(6):686–711.
- Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.
- van Eemeren FH, Grootendorst R, Henkemans FS, Blair JA, Johnson RH, Krabbe ECW, Plantin C, et al. (1996) *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments* (Lawrence Erlbaum Associates, Inc., Mahwah, NJ).
- van Gog T, Rummel N (2010) Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Ed. Psychol. Rev.* 22(2):155–174.
- Vansteenkiste M, Ryan RM (2013) On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *J. Psychotherapy Integration* 23(3):263–280.
- Venable J, Pries-Heje J, Baskerville R (2012) A comprehensive framework for evaluation in design science research. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7286 (Springer, Berlin, Heidelberg), 423–438.
- Venable J, Pries-Heje J, Baskerville R (2016) FEDS: A framework for evaluation in design science research. *Eur. J. Inform. Systems* 25(1):77–89.
- Venkatesh V, Bala H (2008) Technology acceptance model 3 and a research agenda on interventions. *Decision Sci.* 39(2):273–315.
- Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: Toward a unified view. *MIS Quart.* 27(3):425–478.
- Visser J, Lawrence J, Reed C, Wagemans J, Walton D (2022) Annotating argument schemes. Plantin C, ed. *Argumentation Through Languages and Cultures*, 1st ed. (Springer, Cham, Switzerland), 101–139.
- von Aufschnaiter C, Erduran S, Osborne J, Simon S (2008) Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *J. Res. Sci. Teach.* 45(1):101–131.
- vom Brocke J, Maaß W, Buxmann P, Maedche A, Leimeister JM, Pecht G (2018) Future work and enterprise systems. *Bus. Inform. Systems Engng.* 60(4):357–366.
- Vygotsky LS (1980) *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, Cambridge, MA).
- Walton D, Reed C, Macagno F (2008) *Argumentation Schemes* (Cambridge University Press, Cambridge, MA).
- Wang W, Arya D, Novielli N, Cheng J, Guo JLC (2020) ArguLens: Anatomy of community opinions on usability issues using argumentation models. *Conf. Human Factors Comput. Systems Proc.* (Association for Computing Machinery, New York).
- Wambsganss T, Rietsche R (2019) Toward designing an adaptive argumentation learning tool. *40th Internat. Conf. Inform. Systems, ICIS 2019* (AIS Electronic Library, Atlanta).
- Wambsganss T, Janson A, Leimeister JM (2022b) Enhancing argumentative writing with automated feedback and social comparison nudging. *Comput. Ed.* 191:104644.
- Wambsganss T, Küng T, Matthias S, Leimeister JM (2021) ArgueTutor: An adaptive dialog-based learning system for argumentation skills. *Proc. 2021 CHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York).
- Wambsganss T, Swamy V, Rietsche R, Käser T (2022a) Bias at a second glance: A deep dive into bias for German educational peer-review data modeling. *Proc. 29th Internat. Conf. Comput. Linguistics (Gyeongju, Republic of Korea)*, 1344–1356.
- Wambsganss T, Niklaus C, Söllner M, Handschuh S, Leimeister JM (2020b) A corpus for argumentative writing support in German. *28th Internat. Conf. Comput. Linguistics (Coling) (Barcelona, Spain)*.
- Wambsganss T, Niklaus C, Cetto M, Söllner M, Handschuh S, Leimeister JM (2020a) AL: An adaptive learning support system for argumentation skills. *Proc. 2020 CHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York).
- Weber F, Wambsganss T, Neshaei P, Söllner M (2024) LegalWriter: An intelligent writing support system for structured and persuasive legal case writing for novice law students. *Proc. 2024 CHI Conf. Human Factors Comput. Systems (ACM, New York)*.
- WEF (2018) The future of jobs report 2018. *World Economic Forum* (September 17), <https://www.weforum.org/publications/the-future-of-jobs-report-2018/>.
- Weinberger A, Fischer F (2006) A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Comput. Ed.* 46(1):71–95.
- Weinberger A, Stegmann K, Fischer F (2010) Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Comput. Human Behav.* 26(4):506–515.
- Winkler R, Söllner M, Leimeister JM (2021) Enhancing problem-solving skills with smart personal assistant technology. *Comput. Ed.* 165:104148.
- Xu W, Dainoff MJ, Ge L, Gao Z (2021) From human-computer interaction to human-AI interaction: New challenges and opportunities for enabling human-centered AI, 1–73.
- Yi MY, Davis FD (2003) Developing and validating an observational learning model of computer software training and skill acquisition. *Inform. Systems Res.* 14(2):146–169.
- Zacher H, Frese M (2018) Action regulation theory: Foundations, current knowledge and future directions. *The SAGE Handbook of Industrial, Work & Organizational Psychology: Organizational Psychology*, vol. 2, 2nd ed. (Sage Reference, Thousand Oaks, CA), 122–144.
- Zhang F, Hwa R, Litman DB, Hashemi H (2016) ArgRewrite: A web-based revision assistant for argumentative writings. *Proc. 2016 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Demonstrations* (Association for Computational Linguistics, Stroudsburg, PA), 37–41.
- Zimmerman BJ, Schunk DH (2001) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives* (Routledge, New York).