Towards Assisted Excellence: Designing an AI-Based System for Presentation Slide Evaluation

Alexander Meier[1[0009-0000-6279-8343]], Roman Rietsche[2[0000-0002-6112-1709]], and Ivo Blohm[1[0000-0003-2422-5952]]

[1] University of St. Gallen, Müller Friedbergstr. 8, 9000 St. Gallen, Switzerland
alexander.meier@unisg.ch
ivo.blohm@unig.ch
[2] Bern University of Applied Sciences, Falkenpl. 24, 3012 Bern, Switzerland
roman.rietsche@bfh.ch

Abstract. Creating and disseminating high-quality presentation slides have become a foundation for effective communication in educational, corporate, and scientific domains. This study addresses the challenge of enhancing the quality of user-generated presentation content amidst the vast quantities of existing resources. The study is concerned with an ongoing design science research project that focuses on constructing a nascent design theory for a novel AI-based slide evaluation support system (SESS) that aims to assist users, particularly educators, create high-quality presentation slides. The proposed concept leverages recent developments in Generative Artificial Intelligence (genAI) to analyze multimodal user-generated content. Drawing on signaling theory, user-generated online reviews, and expert interviews, this research aims to contribute to delineating the capabilities of artificial intelligence in digital content evaluation, specifically in assisting users to improve the quality of their presentation slides. For practitioners, we offer a set of generalized design principles and design features for the implementation in the development of an AI-based SESS.
Keywords: Generative AI, Design Science Research, Presentation Slides.

1 Introduction

The internet's pervasiveness and ubiquity across all research and practice domains have spurred novel ways of sourcing and disseminating user-generated content [1]. Platforms leverage this by aggregating a vast amount of content uploaded by individuals, thus providing value [2]. In the current information-rich era, presentation slides have emerged as a crucial tool for communicating "texto visual" information in educational contexts, corporate meetings, and scientific dissemination. Presentation slides play a critical role in enhancing learning with visual aids and offering accessible study materials [3]. Determining and creating high-quality presentations requires a combination of didactic and subject-specific expertise. Although user-generated content significantly offers value, its worth hinges on the content's quality [4]. A 2016 poll revealed that 500 million users deliver roughly 35 million PowerPoint presentations daily [5], with SlideShare.net hosting over 40 million presentations [6]. Despite these platforms offering extensive collections of user-generated content, the quality varies widely. Recent advancements in generative artificial intelligence (genAI) have further lowered the barriers to accessing personalized ready-to-use presentations. However, also the performance of state-of-the-art genAI solutions in producing high-quality presentations remains yet unreliable [7].

The challenge for users, especially educators, is to identify high-quality content among the overwhelming volume of data. Information asymmetries arise among users because of differences in educational backgrounds and experience in creating presentations. Artificial intelligence (AI) can learn and save complex data representations from training materials in pre-trained models, later applying this knowledge to unseen data [8]. In line with Signaling theory to address information asymmetries among users with diverse knowledge backgrounds, AI models, trained on thousands of samples, offer to communicate the incorporated knowledge to a broad audience.

Few scholars have used automated approaches to assess digital presentation slides. Most approaches focus on either the technical machine learning model construction or on facilitating the learning for the receiving crowd [9, 10]. However, one of the main challenges to automatically evaluate the quality remains the unstructured and multimodal content of presentation slides, comprising visual and textual elements.

One potential solution to overcome these challenges is the recent advancements in genAI that opened up new avenues for processing this kind of content [11]. What is still missing is a nascent design theory [12, 13] that guides the development of software tools to support content creators in not only assessing the quality of presentation slides but also suggesting areas for enhancing high-quality presentation slides.

Such a presentation slide evaluation support system (SESS) should empower content creators and platforms to effectively identify, create, or reuse high-quality presentations. We actively employ a design science research (DSR) approach to develop prescriptive knowledge, drawing from the foundational frameworks established by Hevner et al. and Peffers et al. [14, 15]. Our primary objective is thus to address the following research question (RQ):

What nascent design theory should guide the development of an AI-based SESS that guides the creation of successful presentation slides?

In this paper, we present design principles (DPs) based on literature, expert insights, and online reviews, which are then transformed into design features (DFs) for an AI-based prototype. We also detail an experimental setup for the first instantiation evaluation. Our contribution includes design principles aiding in the development of automation and AI-supported information evaluation tools already existing in other contexts, such as scholarly publishing [16], healthcare, and data analytics [17]. Additionally, our work aims to anchor user-centered design in Signaling research, drawing on extensive user review analysis.

2 Foundations and Related Work

2.1 Quality in Presentation Slides

While information quality is researched broadly, especially in information retrieval, information quality, specifically in presentation slides, received little attention [18]. Recently, the first approaches emerged to assess slide-related quality, such as the overall presentation skills of students. Luzardo et al. [9], therefore, suggested the use of straightforward features like word count, images, tables, and audio qualities in a pitch, employing a machine learning model to analyze 448 presentations. They classified students' binary as high or low performers. Incorporating this, they reached an accuracy of 65% in evaluating slides and 69% in audio features. Similarly, [10] narrowed their focus to the examination of individual slides. They utilized a neural network approach to differentiate between slides that were novice in design and those that were expertly crafted. This differentiation was based on an analysis of both visual and structural elements inherent in the slides.
Finally, Kim et al. [19] enriched the discourse on presentation slide quality with a comprehensive qualitative study. This involved observing users as they evaluated slides, coding these evaluations, and distilling them into a taxonomy of quality criteria. The resulting taxonomy encompasses four categories: *intrinsic*, *representational*, *contextual*, and *reputational* quality. *Intrinsic* quality focuses on the truthfulness and clarity of objectives in the slides. *Representational* quality assesses the slides' understandability, visual appeal, and navigability. *Contextual* quality covers the completeness, informativeness, timeliness, and task suitability of the information presented. Finally, *reputational* quality reflects the credibility of the slide’s source, often linked to the author's reputation.

2.2 Signaling as Kernel Theory

Research has shown that Signaling Theory supports our underlying hypothesis that individual and personal feedback on an instructor’s ability to design and build slides appropriately influences the instructor to improve their presentation slides [20]. Traditionally, signaling theory is used to explain online market behavior or situations in the job market of multiple parties [21, 22]. The theory constitutes that an agent (sender) credibly conveys some information to a principal (receiver). For example, online product reviewers transmit information to potential buyers through content-related and reviewer-specific signals, influencing review helpfulness and purchase decision-making [23]. The theory has already been applied to other technology domains and poses to be extended to AI-based slide evaluation. [24].
Content creators use presentation slides to convey information to targeted consumers. Also, AI-generated feedback that is observable through a user interface (UI) can be contextualized in Signaling theory. This is also what we will focus on: the AI feedback features are perceptible by the user and the content creator, which results in a user-centric approach for content creators. An AI model can manipulate quality feedback features, such as the projected number of views a presentation could attract on slide-sharing platforms or the text-to-picture ratio, which are observable by the user and,

therefore, qualify as signals. According to [25], information asymmetries arise between those who hold information and those who could potentially make better decisions if they had it, and a specialized trained AI model inherently possesses quality-related information exceeding that of the receiver. It is suggested that the presentation and manipulation of information significantly impact users' decisions on improving, publishing, or selecting presentations, ruling out asymmetry.

3 Design Science Research Approach

To develop a design theory, we deploy a DSR approach. DSR in the field of information systems offers a well-recognized and systematic paradigm for the creation of artifacts expanding on human capabilities [14]. Our study is explicitly concerned with developing a nascent design theory for an AI-based SESS [12, 14]. Our resulting design principles leverage large language models, deep learning algorithms, and a suitable user interface. This embodies a dual function: guiding the initial design and informing subsequent artifact development [26, 27]. Our nascent design theory should integrate requirements (relevance), development practices, and distinctive solutions supported by testable propositions and grounded in foundational theories (rigor) [28].
Concerning the research process design, DSR literature proposes that artifacts advance in an iterative pattern to constantly reflect, build, and refine [14]. We adopt the well-established DSR cycle recommended by Peffers et al. [15] and plan to conduct at least three iterations. This brings together the typical six phases seen in existing DSR literature [29, 30]. First, we started by identifying relevance as outlined in Section 1 (problem identification & motivation phase). We accordingly collected and clustered requirements from primary and secondary data sources. This twofold data collection comprised analyzing user reviews of fully automated AI slide generators, which was supplemented by four expert interviews with re- and upskilling instructors (objectives of a solution phase). This laid the ground for our DPs that were used to address these DRs (design and development phase). Ultimately, we demonstrate our first implementation of a fully functional user interface (demonstration phase). For evaluating (evaluation phase) our instantiated prototype, we are proposing a summative experimental evaluation.

4 Designing AI-based Slide Evaluation Software to Support Content Creator

4.1 Developing Design Requirements

As outlined above, the approach for eliciting DRs is twofold. First, we started by content analyzing online user reviews based on the methods [30] and [31], which they named "crowdsourcing requirements" and have been used on amazon.com and other sites [31]. The *Data Preparation Phase* started with collecting user reviews on Google Workspace Marketplace and Trustpilot. Google Workspace Marketplace is a digital storefront for web-based applications. It facilitates organizational efficiency and offers primarily productivity tools [32]. Trustpilot, in contrast, is an online review community that aims to promote transparency and trust between consumers and businesses. Both allow customers to share their experiences and feedback on products, services, or applications [33]. A total of 600 reviews of AI presentation slide generator apps such as Beautiful.AI, Tome, Decktopus, and SlidesAI.io were collected in January 2024 using a self-programmed Python web crawler. The collected data entries included the complete written text, a pseudonymized user name, a star rating, the platform from which reviews originated, and the app.

Table 1. Categories in user feedback based on themes adapted from Pagano et al. [34]

Theme	Category	Definition	Relative Frequency
Requirements	Feature request	Request for additional features	4.7%
	Content request	Seeks inclusion of new content	3.0%
	Improvement request	Suggests application enhancements	4.4%
User experience	Shortcoming	Specific user dissatisfaction aspect	36.0%
	Helpfulness	Instances of application efficacy	47.0%
	Feature Information	Details on specific features or user interface	7.8%

During the *Content Analysis,* the reviews were coded and checked to ensure they were sufficiently descriptive and meaningful. Reviews like “Terrible” or “All good!!!” were excluded, reducing the number of reviews to N=400. Based on the topics identified in App Store user reviews by Pagano et al. [34], we used the coding categories belonging to the theme requirements and user experience, as shown in Table 1. One, more, or no coding categories can be assigned to one review. In the following, we mapped the most commonly observed user feedback categories onto DRs, applying a triage as not all categories could be considered.

Users who identified shortcomings have been quite harsh in their choice of words (“It does NOT save you time; it is not capable enough to draft you something remotely intelligent.”). A lot of users expressed a flaw in generating and conceptualizing complex presentation topics and the technical stability of the tool (*Shortcoming*). This concludes with the first DR: *Technical reliability and stability*. The users further addressed recurringly either the lack or they praised initial approaches in saving time to build slide decks “The app's ability to generate visually stunning slides with minimal effort on my part has truly revolutionized my workflow” or “In overall, it can reduce only about 20% of our efforts. But not the level which we expect.” (*Helpfulness*). This resonates also with our understanding of the DR, which we collected from advanced studies educators through four semi-structured interviews, all conducted in November 2023 and guided by the expert interview methodology proposed by Gläser [35]. In the interviews, educators uniformly expressed a disinclination toward automated, ready-made educational content generation. Instead, a consensus emerged around the desirability of a tool that facilitates a collaborative approach in the creation of educational materials, coupled with the capability to infuse individualized elements. Central to the discussions among potential users and experienced educators was a salient need: navigating the complexity inherent in the development of visually compelling slides while concurrently ensuring alignment with the user's distinct and complex workflow. This requirement renders the slide creation process time-intensive, necessitating advanced proficiency in diverse areas, including an acute understanding of the target audience and a comprehensive grasp of both textual content and design elements. This requirement resonates with the concept of Signaling, as outlined in Section 2.2, which focuses on mitigating information asymmetries. This observation concludes in DRs within our nascent Design Theory: *promote a knowledge discovery process in slide design*. Not only is the knowledge needed, but the process is also time-consuming; the objective is to Increase time efficiency in creating high-quality slides and fostering and maintaining user-driven creative design expression while maintaining an objective evaluation that *adaptively evolves* over time.

4.2 Deriving Design Principles, Design Features, and a First Instantiation

The approach of [27] guided us to conceptualize the DPs (see Figure 1) with corresponding DFs to tackle the DRs. We specified the DPs using the actors, context, mechanism, and rationale comprising the schema of Gregor et al. [36]. The SESS utilizes an evaluation dashboard as its mechanism, providing instant feedback on design and content. The rationale is that instant feedback allows users to rapidly identify and implement design improvements concluding into DP1. This is powered by a historical database from slide-sharing platforms for training its AI models. In this data-rich environment, the system can analyze successful slide design patterns. Such a vast dataset provides reliable benchmarks, guiding users to improve based on proven standards (DP2) and transparent evaluation criteria. It articulates the rationale behind each metric to users, fostering explainability and providing insights for targeted slide enhancement (DP3). Furthermore, the SESS streamlines presentation narrative structuring (DP4) and refines its AI model through user feedback (DP5).

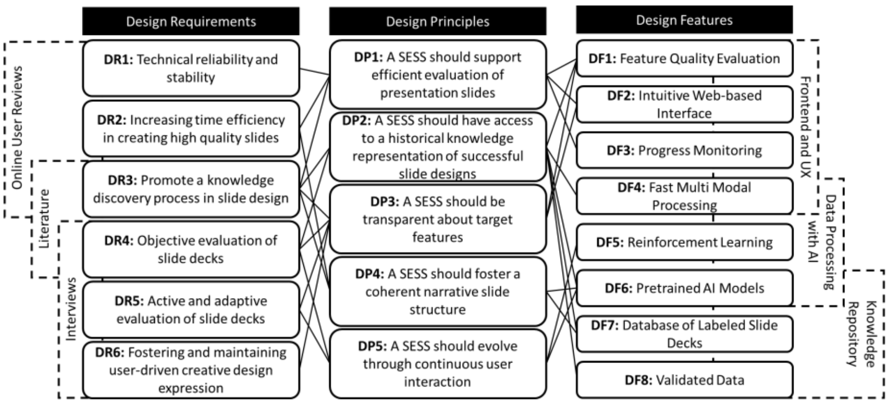

Fig. 1. Conceptualization of Design Requirements, Principles, and Features

We instantiated the first version (see Figure 2) of such a system, as described by DF1-DF8; it requires three core components. A frontend concerned with delivering a seamless user experience (UX) and communicating the evaluation, a backend with multimodal processing techniques and reinforced AI models, and a knowledge repository exhibiting validated quality properties such as the number of public views on slide sharing platforms and best practice slide deck data that has been annotated by experts.

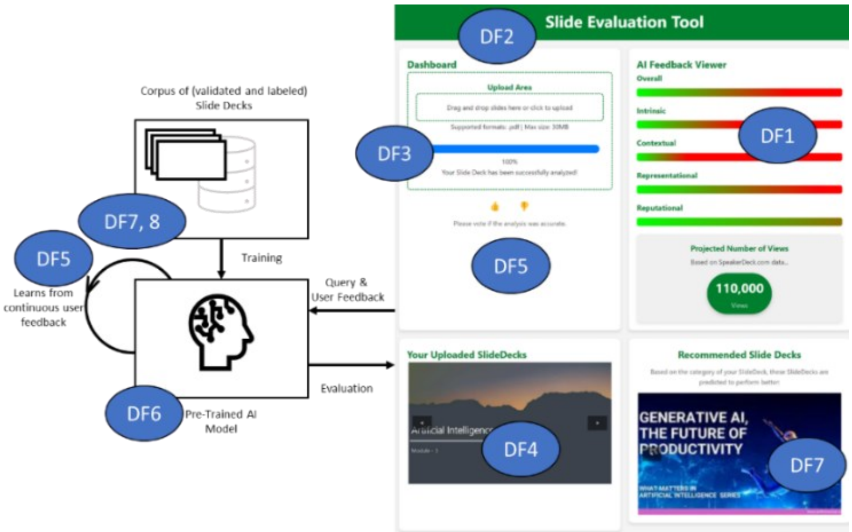

Fig. 2. Backend and Frontend Architecture Including Design Features

4.3 Expected Evaluation

We are planning to run lab and field experiments to evaluate the SESS. Study 1 focuses on a lab experiment that is intended to take place in an ex-post evaluation. We hypothesize that our DPs and the SESS instantiated DFs address the previously defined DRs. Thus, the central hypothesis is that AI-generated evaluation feedback increases the performance in the creation of individuals and, as suggested by Signaling theory (see section 2.2), mitigates deficiencies in slide design knowledge. The first measure incorporates a self-estimation adapted from [37], where participants introspectively evaluate their slide design knowledge on a Likert-type item scale. This introspective evaluation of slide design knowledge is supplemented by a task classifying three pre-assessed slide decks by experts into "low", "fair", and "high" quality, along the four categories adapted from the taxonomy from Kim et al. [19] and the overall quality. Each correctly classified category per slide deck awards one point, yielding a maximum number of 12. These control measures benchmark the design competencies of our participants before the intervention. The study comprises a treatment group, which is assisted by the SESS. The group serves as a fundamental baseline for elucidating the enhancements potentially ascribed to the SESS. The control group, in contrast, solely relies on itself without the provision of the SESS.

Each participant is randomly assigned a short slide deck, previously annotated as underperforming in the categories adapted from Kim et al. [19], and the number of views on SlideShare.net. Both groups are asked to improve the slide deck. The time each participant requires to enhance their slides, both with and without the aid of SESS, is measured, providing insights into potential efficiency gains. Subsequently, in a post-evaluation, the expert panel reconvenes to reassess the enhanced presentation slides by the participants, employing again the categories and dimensions adapted from Kim et al. [19]. This aims to quantify the tangible enhancements in slide design attributable to the intervention.

Ultimately, in a post-survey on fostering personal creative design expression, we operationalize measures on the participants' self-reported creativity [37], enjoyment [38], involvement, self-efficacy, and creative autonomy [39] during the task. Additionally, we ask only the participants in the treatment group after using the SESS for adoption intentions [40], perceived value [41], and perceived usefulness [42].

5 Next Steps, Limitations, and Expected Contribution

Addressing the limitations of this study opens avenues for future research. First, we aim to rigorously evaluate and refine the initial SESS implementation based on the outlined experimental framework. Second, we intend to enhance the SESS by specifically tailoring it for continuing education before broadening its application to domains such as pitch and consulting decks, which have yet to be explored. Third, the current SESS architecture, already implemented relying on a combination of a single-layer neural network, multiple machine learning algorithms like Random Forest, and a generative AI model, may not represent the optimal configuration and are subject to a later iteration.

In summary, our work contributes to the existing body of literature on knowledge evaluation and feedback, providing operational principles [13] to assist practitioners in the deployment and implementation of AI-based knowledge evaluation methods.

Acknowledgments. We acknowledge Julia Logo, Moritz Schmid, Heinrich von Waldersee, and Yaku Colque for permission to use the interview data.

This work was supported by Innosuisse and is part of the flagship project SCESC, PFFS-21-29.

References

1. Krumm J, Davies N, Narayanaswami C (2008) User-Generated Content. IEEE Pervasive Computing 7:10–11. https://doi.org/10.1109/MPRV.2008.85

2. Thelwall M, Kousha K (2016) SlideShare Presentations, Citations, Users and Trends: A Professional Site with Academic and Educational Uses. Journal of the Association for Information Science and Technology 68:. https://doi.org/10.1002/asi.23815

3. Chávez Herting D, Cladellas Pros R, Castelló Tarrida A (2020) Patterns of PowerPoint Use in Higher Education: a Comparison between the Natural, Medical, and Social Sciences. Innov High Educ 45:65–80. https://doi.org/10.1007/s10755-019-09488-4

4. Kim C, Jin M-H, Kim J, Shin N (2012) User perception of the quality, value, and utility of user-generated content. Journal of Electronic Commerce Research 13:305

5. polleverywhere (2016) 10 little-known facts about PowerPoint. https://blog.polleverywhere.com/powerpoint-infographic. Accessed 16 Nov 2023

6. Ha A (2020) Scribd acquires SlideShare from LinkedIn. In: TechCrunch. https://techcrunch.com/2020/08/11/scribd-acquires-slideshare/. Accessed 31 Oct 2023

7. Guo Y, Zhang Z, Liang Y, et al (2023) PPTC Benchmark: Evaluating Large Language Models for PowerPoint Task Completion

8. Deng L, Liu Y (2018) Deep Learning in Natural Language Processing. Springer, Singapore

9. Luzardo G, Guamán B, Chiluiza K, et al (2014) Estimation of Presentations Skills Based on Slides and Audio Features. In: Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge. ACM, Istanbul Turkey, pp 37–44

10. Yi S, Matsugami J, Yamasaki T (2022) Assessment system of presentation slide design using visual and structural features. IEICE TRANSACTIONS on Information and Systems 105:587–596

11. Lee DW, Ahuja C, Liang PP, et al (2023) Lecture Presentations Multimodal Dataset: Towards Understanding Multimodality in Educational Videos. pp 20087–20098

12. Jones D, Gregor S (2007) The Anatomy of a Design Theory. Journal of the Association for Information Systems 8:. https://doi.org/10.17705/1jais.00129

13. Gregor S, Hevner A (2013) Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37:337–356. https://doi.org/10.25300/MISQ/2013/37.2.01

14. Hevner AR, March ST, Park J, Ram S (2004) Design Science in Information Systems Research. MIS Quarterly 28:75–105. https://doi.org/10.2307/25148625

15. Peffers K, Tuunanen T, Rothenberger M, Chatterjee S (2007) A design science research methodology for information systems research. Journal of Management Information Systems 24:45–77

16. Schwartz D, Te'eni D (2024) AI for Knowledge Creation, Curation, and Consumption in Context. Journal of the Association for Information Systems 25:37–47. https://doi.org/10.17705/1jais.00862

17. Pezoulas V, Kourou KD, Kalatzis F, et al (2019) Medical data quality assessment: On the development of an automated framework for medical data curation. Computers in biology and medicine 107:270–283. https://doi.org/10.1016/j.compbiomed.2019.03.001

18. Kim S, Jung W, Han K, et al (2014) Quality-Based Automatic Classification for Presentation Slides. In: De Rijke M, Kenter T, De Vries AP, et al (eds) Advances in Information Retrieval. Springer International Publishing, Cham, pp 638–643

19. Kim S, Lee J-G, Yi MY (2017) Developing information quality assessment framework of presentation slides. Journal of Information Science 43:742–768. https://doi.org/10.1177/0165551516661917

20. Dinh Tho N (2017) Using signals to evaluate the teaching quality of MBA faculty members: fsQCA and SEM findings. Education + Training 59:292–304. https://doi.org/10.1108/ET-03-2016-0060

21. Dimoka A, Hong Y, Pavlou PA (2012) On Product Uncertainty in Online Markets: Theory and Evidence. MIS Quarterly 36:395–426. https://doi.org/10.2307/41703461

22. Spence M (1973) Job Market Signaling. The Quarterly Journal of Economics 87:355–374. https://doi.org/10.2307/1882010

23. Siering M, Muntermann J, Rajagopalan B (2018) Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. Decision Support Systems 108:1–12. https://doi.org/10.1016/j.dss.2018.01.004

24. Guo X, Deng H, Zhang S, Chen G (2020) Signals of Competence and Warmth on E-Commerce Platforms. Data and Information Management 4:81–93. https://doi.org/10.2478/dim-2020-0010

25. Connelly BL, Certo ST, Ireland RD, Reutzel CR (2011) Signaling Theory: A Review and Assessment. Journal of Management 37:39–67. https://doi.org/10.1177/0149206310388419

26. Piirainen KA, Briggs RO (2011) Design Theory in Practice – Making Design Science Research More Transparent. In: Jain H, Sinha AP, Vitharana P (eds) Service-Oriented Perspectives in Design Science Research. Springer, Berlin, Heidelberg, pp 47–61

27. Meth H, Mueller B, Maedche A (2015) Designing a Requirement Mining System. Journal of the Association for Information Systems 16:. https://doi.org/10.17705/1jais.00408

28. Markus ML, Majchrzak A, Gasser L (2002) A Design Theory for Systems That Support Emergent Knowledge Processes. MIS Quarterly 26:179–212

29. Kuechler W, Vaishnavi V (2008) On theory development in design science research: anatomy of a research project. EJIS 17:489–504

30. Walls JG, G. J, Widmeyer, et al (1992) Building an Information System Design Theory for Vigilant EIS. Information Systems Research 3:36. https://doi.org/10.1287/isre.3.1.36

31. Mudambi SM, Schuff D (2010) Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. MIS Quarterly 34:185–200. https://doi.org/10.2307/20721420

32. Google Workspace Marketplace. In: Google for Developers. https://developers.google.com/workspace/marketplace. Accessed 17 Jan 2024

33. What's the purpose of Trustpilot? In: Trustpilot Help Center. https://support.trustpilot.com/hc/en-us/articles/222997828--What-s-the-purpose-of-Trustpilot. Accessed 17 Jan 2024

34. Pagano D, Maalej W (2013) User Feedback in the AppStore: An Empirical Study

35. Gläser J, Laudel G (2010) Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. Experteninterviews und qualitative Inhaltsanalyse

36. Gregor S, Kruse LC, Seidel S (2020) Research Perspectives: The Anatomy of a Design Principle. Journal of the Association for Information Systems 21:. https://doi.org/10.17705/1jais.00649

37. Bruner GC (2015) Marketing scales handbook ; 7. Multi-item measures for consumer insight research / Gordon C. Bruner. GCBII Productions, Fort Worth, Tex

38. Agarwal R, Karahanna E (2000) Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS Quarterly 24:665–694

39. Bruner II GC (2012) Marketing scales handbook. GCBII Productions

40. Davis FD, Bagozzi RP, Warshaw PR (1989) User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35:982–1003. https://doi.org/10.1287/mnsc.35.8.982

41. Sirdeshmukh D, Singh J, Sabol B (2002) Consumer Trust, Value, and Loyalty in Relational Exchanges. Journal of Marketing 66:15–37

42. Kim H-W, Chan HC, Gupta S (2007) Value-based Adoption of Mobile Internet: An empirical investigation. Decision Support Systems 43:111–126. https://doi.org/10.1016/j.dss.2005.05.009