MethodsX 13 (2024) 102950

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Usability testing of a conceptual model through retrospective cross-case analysis ☆,☆☆

John Paul Mynott [a], Klara Kager [b,*]

[a] School of Education, University of Aberdeen, Scotland
[b] Department for Research on Teaching and Educational Interventons, University of Potsdam, Germany

ARTICLE INFO

Method name:
Usability Testing Conceptual Models

Keywords:
Usability test
Conceptual model
Cross-case analysis
Retrospective analysis
Qualitative evaluation

ABSTRACT

Qualitative conceptual models are commonly used in the scientific literature to make complex phenomena easier to understand. However, the effectiveness and usability of conceptual models to serve as analytical tools is rarely explored and tested, and there is a lack of guidelines for such analyses. This paper adapts and combines the methods of usability testing and cross-case analysis to describe a systematic protocol to facilitate the qualitative evaluation of conceptual models. Usability testing is an established method for identifying problems or shortcomings within a product and for assessing different dimensions of product usability: suitability, accessibility, relevance, and integrity. Cross-case analysis, on the other hand, is a qualitative research method for systematically comparing information from individual case studies and identifying commonalities and patterns that apply across cases. Taken together, these methodological approaches provide a structured way of retrospectively applying a conceptual model to existing literature and thereby evaluating its effectiveness in meeting its intended purpose.

- We show how researchers can prepare, conduct, and synthesise the results of a usability test of a conceptual model.
- We provide recommendations for the practical implementation of each step.
- We outline the benefits, limitations, and ethical considerations that researchers should be aware of.

Specifications table

Subject area:	Economics and Finance
More specific subject area:	*Professional Development in Education*
Name of your method:	*Usability Testing Conceptual Models*
Name and reference of original method:	*Usability Testing, Cross-Case Analysis*
Resource availability:	*None*

☆ **Related research article:** Authors. (2024). Applying a conceptual model to understand quality in Lesson Study: a cross-case usability test. [Manuscript submitted for publication]
☆☆ **For a published article:** None.

* Corresponding author.
E-mail address: Klara.kager@uni-potsdam.de (K. Kager).

https://doi.org/10.1016/j.mex.2024.102950
Received 18 June 2024; Accepted 4 September 2024
Available online 5 September 2024

Background

Conceptual models in social sciences aim to provide a comprehensive picture of a complex phenomenon or process, helping us to better understand a problem and how variables interact with it [6]. This makes conceptual models the basis for describing objects, processes, and outcomes, and for theorising how they influence each other [12]. By their very nature, conceptual models are incomplete [16] and should be iteratively applied, tested, and improved [6].

Conceptual models are typically used prospectively to guide empirical research, and are improved based on new findings from their application. Yet, models can also be used retrospectively as an analytical tool to understand and re-contextualise existing research [25]. While there are some examples of studies that use conceptual models to categorise the output of a research field (e.g., [5]), there are few examples of conceptual models being explored and tested as an analytical tool through such a qualitative application. Importantly, there is a lack of guidelines on how to undertake such an analysis. However, using a conceptual model as an analytical tool could reveal critical initial insights into the completeness and usability of a model, particularly if a model is new, represents an innovation, or has not been yet been applied in the field. This paper addresses this gap in the current research by setting out a method of a usability test using cross-case analysis to retrospectively explore existing research as a way of testing a conceptual model.

In itself, the retrospective analysis of existing research is a common methodological approach in the qualitative social science literature, most often used to systematically synthesise case studies on a given phenomenon [4]. Such a cross-case analysis can be a necessary step for a research field to accumulate knowledge that can be generalised beyond a single case [4]. Khan and VanWynsberghe [15] therefore refer to cross-case analysis as "a mechanism for mining case studies so that knowledge from cases can be put into service for broader purposes" (p. 1). Although this approach is guided by protocols and is considered to be highly systematic and rigorous [4], it does not appear to have been used as a means of evaluating the effectiveness or usability of the models, and by extension the protocols, guiding the synthesis itself.

The concepts of effectiveness, evaluation, and validation are sometimes considered controversial and borrowed from quantitative research by qualitative researchers [3]. Some even argue against attempts to develop standardised measures or protocols for qualitative research to prevent the quantification of qualitative research paradigms [2,3,28]. Nevertheless, there remains a need to address and explore quality and effectiveness of tools used in qualitative research, such as conceptual models [2,18]. The field of usability testing provides an established method for testing the effectiveness and usability of products [10]. If we consider a conceptual model as a tool that researchers – the users – engage with, usability testing is a viable way to identify issues or shortcomings within the conceptual model [17] and to better understand their complexity [9].

Method details

Combining the methodological approaches of cross-case analysis and usability testing, we developed a systematic and replicable way of testing the usability of a conceptual model as an analytical tool. As usability relates to how easy it is for a user to use a product, it is important that a usability test uses an evaluation method that is appropriate for both the test and the users [14]. Usability testing has implemented a diverse range of evaluative methods [14] but our use of cross-case analysis provides the user with ability to compare existing research, through the conceptual model being tested, to see if the model is effective as an analytical tool. In a first step, we define the dimensions of usability testing and relate them to the testing of conceptual models. In a second step, we adapt the methodological approach of cross-case analysis to develop a systematic protocol so that usability testing can be applied to a conceptual model. We developed this approach in order to test a conceptual model for teacher learning through continuous professional development within the field of educational research [13]. We will use examples from this co-submitted paper to contextualise aspects of this methods paper and to illustrate or clarify aspects of our discussion.

Method basics: usability testing

Usability testing serves as both a quantitative and qualitative evaluative method that draws attention to problems with a given product [17], but can also highlight positive aspects related to effectiveness, efficiency, and user satisfaction [1]. While usability testing is a common approach across research fields, there is disagreement about which dimensions should be assessed and how we should refer to them. As Hankiewicz [9] notes, researchers use similar yet different definitions and also tend to focus on the aesthetic presentation of these terms. For example, Morville [19] presents their terminology in a honeycomb diagram, while Quesenbery [24] speaks of the 5E's of usability testing: effective, easy to learn, engaging, efficiency, error tolerant. These guidelines can be visually or linguistically appealing, yet the focus on aesthetics means that the field now has multiple terms that appear similar but diverge in meaning.

The challenge in conducting qualitative usability testing on a conceptual model is therefore to first define the dimensions and terminology linked to the test. To map the divergent terminology, we used the snowball principle. Consulting recent overview texts from both the qualitative validation and usability literature [3,9], we identified prominent papers, in their citations from both fields that discuss and define terminology related to the dimensions of usability. We then mapped the terminology against each other (Table 1) using definitions given by the respective authors. As Table 1 shows, both the dimensions and terminology differ between articles. For example, Prussak and Hankiewicz [23] and Nielsen [21] define errors within the concept of suitability for a given task, whereas Quesenbery's (2004) definition is more overtly linked to the relevance of the test. Overall, the usability testing literature tends to focus on the appearance, suitability, and ease of use as primary factors, with integrity only addressed in the recommendations of Morville [19] and Prussak and Hankiewicz's [23]. This suggests that, especially when testing the usability of websites and/or

Table 1
Mapping the diverging terminology related to validation and usability.

	Validation Literature		Usability Literature			
	Whittlemore et al. [29]	Tracy [28]	Morville [19]	Quesenbery [24]	Prussak & Hankiewicz [23]	Neilsen [21]
Suitability	-authenticity -explicitness -thoroughness	-contribution	-useful -valuable	-effective	-error tolerance -adequateness -usefulness	-errors -satisfaction
Relevance	-criticality -congruence	-worthy -rigour -resonance -coherence	-desirable -findable	-error tolerant	-self-descriptiveness -comprehension	
Accessibility	-vividness		-accessible -usable	-easy to learn -engaging -efficiency	-aesthetics -accessibility -ease learning of use -using fastness -easiness of use	-learnability -efficiency -memorability
Integrity	-credibility -integrity -sensitivity	-sincerity -credibility -ethical	-credible		-integrity	

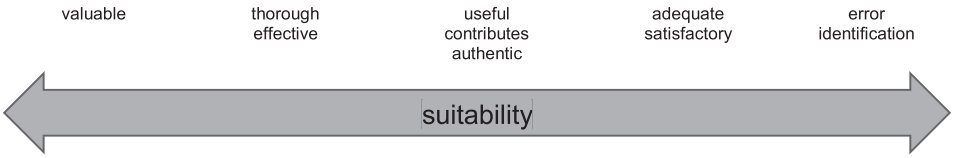

Fig. 1. The continuum of the dimension of suitability.

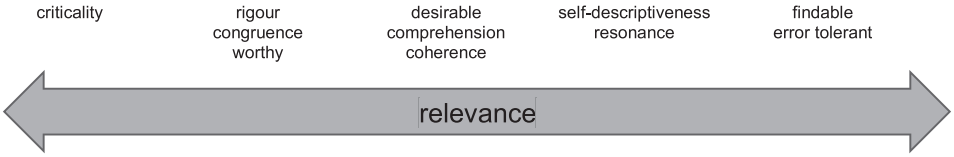

Fig. 2. The continuum of the dimensions of relevance.

programs featuring distinct user interfaces, user experience tends to take precedence over product integrity. Conversely, there is a notable focus in the validation literature on both the relevance and integrity of the research. This emphasis likely stems from the goal of verifying observed and defined phenomena [28] and managing external pressures related to the measurability of qualitative research [2]. Grouping the divergent terminology into four general codes allows us to make a connection between these terms and to define four relevant dimensions for the usability testing of conceptual models: Suitability, Accessibility, Relevance, and Integrity (Table 1).

Suitability

Suitability refers to the effectiveness [24] or value [19] of a model in supporting the user's thinking, analysis, or understanding of a given phenomenon. If the divergent terminology is considered as a continuum from error identification through usefulness to value, it is possible to see that all the divergent terms relate to whether or not a model is suitable for a given purpose (Fig. 1).

The suitability dimension is typically assessed through user satisfaction or error rates in the usability literature [21]. However, this does not take into account the value dimensions [19,29] and given the purpose of conceptual models to facilitate the generation of new insights into a complex situation [6], a conceptual model can be considered suitable if it effectively supports users in describing a phenomenon and/ or make new connections that would have remained hidden without the abstraction provided by the model. Therefore, we define the suitability of a conceptual model as its ability to support the user in thinking about, analysing, or exploring a given phenomenon, thereby enhancing the user's understanding of a complex situation.

Relevance

Relevance refers to the extent to which a product satisfies the specific needs and requirements of its intended audience and is congruent with its purpose [29]. Unlike the broader dimension of suitability for a task, this dimension therefore assesses whether a conceptual model, when used as a tool, is fit for its specific purpose [9]. The variation in terminology (Fig. 2) means that relevance ranges from whether the model can describe a given phenomenon through to whether it can act as an analytical tool to enhance the

Fig. 3. The continuum of the dimensions of accessibility.

users understanding of that phenomenon. This means that in order to assess the relevance of a conceptual model, the user will need to set specific criteria for the model's purpose and then analyse if the model helps achieve these criteria.

Accessibility

The dimension of accessibility refers to the ease and satisfaction of using and navigating a product [23]. While a range of terminology is used to refer to these aspects (Table 1), Fig. 3 shows that these terms are closer in meaning than in the two previous dimensions, with many of the terms related to how easy a model is to use, learn or remember.

According to Quesenbery [24], a user's ability to comprehend a product or learn how to operate it is crucial in preventing disengagement. A user-friendly design is therefore not directly related to the suitability of a product, but it may predict its success. An overly complex product, even if it meets other usability criteria, might disengage its audience. An accessible conceptual model therefore needs to use a clear terminology and layout [29] that are easy for its intended users to understand and navigate.

Integrity

The least frequently mentioned dimension is integrity, which refers to the overall credibility [19] and sincerity [28] of a product. This dimension is linked to the genuineness of a product and therefore builds on its suitability, accessibility, and relevance. Integrity thus differs from the previous three dimensions in that the integrity of a product would be derived from the usability outcomes of the previous three dimensions. The integrity of a conceptual model can be seen as a summary of whether it is found to be suitable, accessible, and relevant through a series of usability tests, and is therefore a dimension of usability that is important for users to consider, but needs to be evaluated over time.

Testing protocol: combining usability testing and cross-case analysis

In the following, we will outline a testing protocol based on the setup of a usability test; however, rather than having users engage with and evaluate a product, they engage with a conceptual model and use it as an analytical tool to conduct a cross-case analysis. In other words, our approach shifts the focus from evaluating a product's usability through questionnaires or focus groups to assessing a model's completeness and relevance by applying it to existing research studies.

The basic setup of a usability test involves users engaging with a product and continuously thinking aloud [7], while an observer listens and records notes [10]. Several protocols have been developed to structure this setup (e.g., [10,27]), which include preparatory steps, such as design or contextualisation, the implementation or use itself, as well as aspects of analysis, evaluation, and learning. Given the inconsistent use of terminology to refer to these steps within the field of usability testing, and the fact that our procedure includes the retrospective analysis of literature, it is helpful to move away from the usability literature and use established terminology from qualitative research methods that will be familiar to a wide range of researchers in social sciences. Drawing on methodological recommendations by Cruzes et al. [4] and Khan and VanWynsberghe [15], we propose a protocol consisting of three main phases that include several steps.

Phase 1: preparing

The goal of the first stage is to prepare the subsequent use of the conceptual model by (1) defining the scope of the usability test and (2) defining the sample. The first step in defining the scope of the usability test is to contextualise the conceptual model and its development. This includes positioning the model within its respective domain, addressing why the model was created, and clarifying any additional terminology or frameworks necessary to understand the model. This contextualisation is linked to establishing the criteria that indicate whether the given model is effective and useful. For example, the purpose of the model tested in the co-submitted paper was to address specific gaps in previous models on teacher professional development, thus providing a way to study the same professional development approach across cultural contexts and educational settings. One criterion to test the model's relevance would therefore be its cross-cultural applicability [13]. As Whittlemore et al. [29] and Tracy [28] argue, formulating specific criteria that the model is intended to meet brings explicitness and rigour to the usability test.

Defining the scope of the usability test is a necessary first step in order to select the sample articles for the analysis. Selecting the sample includes the steps of (1) establishing criteria for sample selection and (2) applying purposive sampling [22]. These steps are linked to the thoroughness [29] and accuracy [3] in the usability testing literature. The first step is to outline the criteria that mark articles as suitable for the analysis. These criteria need to be informed by the purpose of the usability testing [29], which may

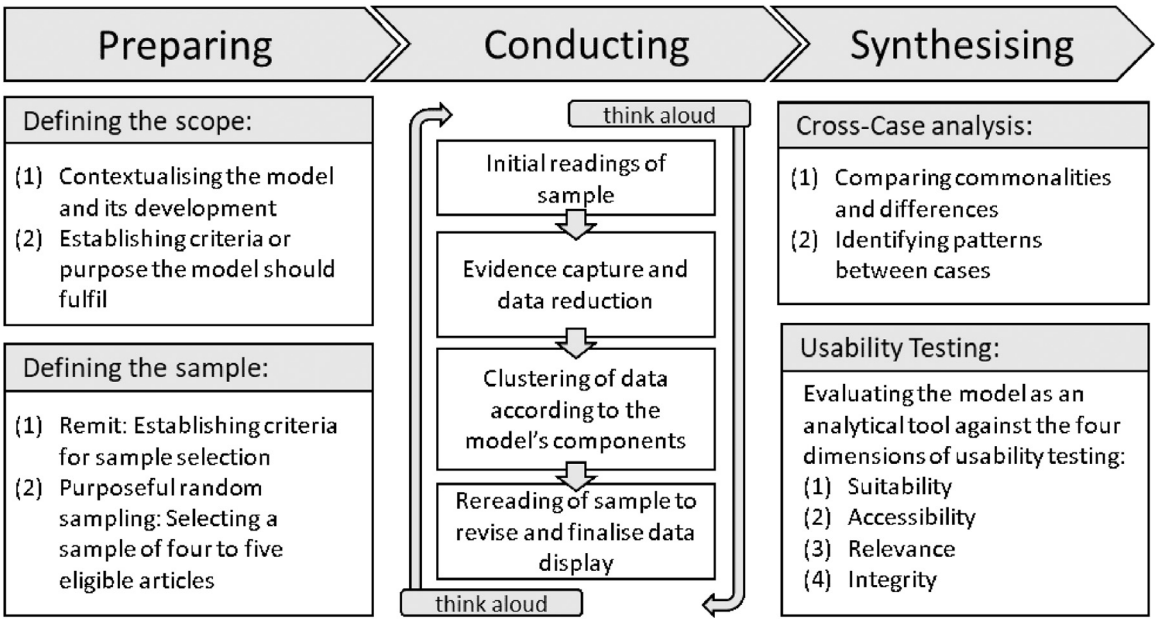

Fig. 4. Protocol for usability testing using cross-case analysis.

require literature from a particular geographical area, time frame, or using a particular theoretical framework. For example, in our co-submitted research [13], the purpose of the conceptual model required us to select an international sample that included cases from various cultural contexts. At the same time, our research questions required us to limit this sample to cases in which the authors had identified specific challenges to the quality of the investigated professional development approach. These criteria determine which research articles are eligible for inclusion in the final sample and which are not.

Once an eligible sample has been identified, four to five articles should be selected from that sample [15]. In usability testing, three to five cases are considered to be in the golden zone [1,20], as this sample size should reveal most of the significant issues with a tool [17,20]. As Barnum [1] argues, larger sample sizes are only useful for summative assessments of usability, not formative and qualitative assessments such as those outlined in our protocol. While the selection of articles should be random, it also needs to be purposive [22] in order to serve the purpose of evaluation. Our co-submitted research, for instance, aimed to assess a model's cross-cultural applicability [13]. We therefore had to purposefully include articles from different national, cultural, and educational contexts.

Phase 2: conducting

The next phase concerns the iterative analysis of the sample, which serves two purposes: (1) to conduct a cross-case analysis and (2) to assess the usability of the model. The cross-case analysis involves several steps carried out by the users of the model, who simultaneously act as evaluators. In a first step, the users individually read each sample paper multiple times to familiarize themselves with the cases. In a next step, users summarize the information and evidence from each article in order to reduce and collate their findings [4]. The findings are then clustered and categorized according to the themes or component parts of the conceptual model. These steps– the capturing of evidence and clustering of data–should be supported by systematic tools, such as a structured list of the model's components. The goal is to develop an organized data display that will later serve as the basis for drawing conclusion [4]. In a final step, the users reread each article to verify the correctness of the information in the data display.

Importantly, these steps are an iterative process, which is illustrated in Fig. 4 by surrounding the steps of Phase 2 with arrows. The goal of this iterative process is for users to continuously clarify their thinking, address biases, and engage with the relevant sample for a prolonged time [3]. This serves both purposes of Phase 2: it facilitates the cross-case analysis and enables for usability testing. The thinking aloud method is traditionally considered the "gold standard" ([11], p. 107) in usability testing, as it allows for rich descriptions of aspects of the usability of a given product. In practice, this means that users take notes of their impressions and, at certain points in the process, discuss their thinking and findings with each other [10]. These points in time are opportunities to ensure corroboration of information [3] and to reveal tacit thinking about the use of the model [28]. The repeated exchange can reveal positive and negative opinions on aspects of the model [26], which provides insights on the accessibility, relevance, and usability of the model in relation to its intended purpose [10]. Although the literature does not offer recommendations regarding to the number of users – or researchers – who should be involved in this iterative analysis, we recommend a minimum of two in order to facilitate the thinking aloud method.

Phase 3: synthesising

The last phase concerns the synthesis of the findings in order to draw conclusion on (1) the cross-case analysis and (2) the usability testing of the model. Regarding the cross-case analysis, the users now consider the data display and compare the clustered data between cases [4]. The goal is to identify commonalities, differences, and patterns between cases that emerged because of the clustering and re-categorization of the data according to the components of the conceptual model.

Regarding the usability testing, the goal of the final phase is to evaluate the model as an analytical tool against the four dimensions of usability testing: Suitability, Relevance, Accessibility, and Integrity. For example, a model may be considered suitable if it effectively supported the cross-case analysis and enhanced the users' understanding of the phenomenon beyond the single case. In our co-submitted research paper, we aimed to demonstrate the model's suitability by focusing on an aspect that had been relatively underexplored in previous studies: the quality of learning processes as an outcome of professional development. We posed the research question: "What does the use of the conceptual model reveal about the quality component?" To answer this, we condu cted a cross-case analysis of four studies that provided data on learning quality. By using the model to recontextualize and synthesize the findings from each study, we were able to make meaningful comparisons and draw broader conclusions. A model may be considered relevant if it effectively fulfils the criteria for which it was developed. In our co-submitted paper, we tested a model designed to meet the following criteria: (1) to be usable across different national and cultural contexts; (2) to facilitate systematic descriptions of all components of the professional development intervention; (3) to serve as an analytical tool for examining the impact of the intervention; and (4) to facilitate the ongoing and formative use of the intervention [13]. To demonstrate the model's relevance, our cross-case analysis needed to show that the model—used as the protocol for data extraction—(1) could be applied to case studies from different contexts, (2) encompassed all elements and components described in these case studies, (3) generated new knowledge by enabling comparisons between case studies, and (4) captured both short-term and long-term outcomes of the intervention [13]. The conceptual model we tested was also found to be highly relevant enabling a shared language through which we could explore quality in Lesson Study [13]. A model may be considered accessible if the users perceive it as easy to use and quickly become proficient in its application. In our co-submitted paper, for example, we evaluated whether the terminology of the conceptual model was accessible and useful for contextualizing the research studies selected for the cross-case analysis [13]. We found that the model provided a common language and allowed comparison of aspects that were referred to differently in each sample article. We also considered whether the layout of the model needed to be adapted in order to serve as an analytical tool. In our case, we created a simple spreadsheet that allowed data to be extracted from each article and for each component of the model. The terminology of the conceptual model we tested was accessible and helped create shared cognition of the ideas linked to professional development through Lesson Study [13]. Finally, a model may be considered to have integrity if, over time, its application is consistent with the purpose for which it was originally designed. For example, the findings in our co-submitted research on the usability of the conceptual model that we tested are a contribution towards understanding the model's integrity as a way to facilitate coherent and comprehensive descriptions of the examined professional development approach across different cultural and educational settings (Kager & Mynott, 2022). This means that in a single usability test protocol it may be difficult to determine the integrity of a conceptual model, but when multiple tests have been conducted, the collective results can be used to consider the integrity of the model.

Method validation

None.

Limitations

The method presented in this paper offers a systematic and transparent way to explore and evaluate the usability and effectiveness of a conceptual model through its application to existing literature. This approach can help identify potential usability issues or problems with the components of the model, as well as implications for its iterative improvement [1]. In addition, this method suggests a way of applying a common conceptual framework to individual case studies, facilitating their synthesis. Traditionally, cross-case analysis is described as an inductive process in which data from case studies are reduced and categorised in order to generate codes, graphs, or networks [4]. We propose a deductive approach in which a priori categories are derived from a conceptual model to which the data retrieved from the literature are then assigned. This process allowed us to analyse the individual cases in our co-submitted research paper to be conceptualised against a common backdrop, revealing commonalities between cases that would otherwise have remained hidden. The analysis also facilitated the development of a shared terminology, as the model provided an overarching terminology to talk about aspects that were referred to in different ways in the individual studies.

However, it is important to also consider the limitations of this approach. A notable constraint of cross-case analysis is the potential for the oversimplification of the reduced data which may lead to a superficial understanding of a complex phenomenon [4]. In addition, there is a risk for research bias in the selection and interpretation of cases [30]. Khan and VanWynsberghe [15], citing Goetz and LeCompte [8], further caution against the notion of generalising findings from case studies, suggesting instead that efforts should be made to enable comparability between cases. Finally, it is important to acknowledge the ethical implications of retrospectively using published research and revisiting it through the lens of a conceptual model. We therefore suggest the following practical concerns:

- Purposively selecting articles from an eligible sample.
- Assigning the collected data to the components of the model with care and without interpretation, bearing in mind that the original articles may not contain all the elements under considered.
- Considering whether consent from the original authors is needed, for example if the articles are unpublished manuscripts (doctoral theses etc.).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

John Paul Mynott: Conceptualization, Methodology, Visualization, Writing – original draft. **Klara Kager:** Conceptualization, Methodology, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Ethics statements

None.

Acknowledgments

Klara Kager was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF).

Supplementary material and/or additional information [OPTIONAL]

None.

References

[1] C. Barnum, Usability testing essentials: Ready, set… test!. ProQuest Ebook Central. (2010).
[2] J. Cho, A. Trent, Validity in qualitative research revisited, Qual. Res. 6 (3) (2006) 319–340, doi:10.1177/1468794106065006.
[3] J. Cresswell, C. Poth, Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 4th ed., Sage, 2018.
[4] D.S. Cruzes, T. Dybå, P. Runeson, M. Höst, Case studies synthesis: a thematic, cross-case, and narrative synthesis worked example, Empir. Softw. Eng. 20 (6) (2015) 1634–1665, doi:10.1007/s10664-014-9326-8.
[5] L. Darragh, Identity research in mathematics education, Educ. Stud. Math. 93 (1) (2016) 19–33, doi:10.1007/s10649-016-9696-5.
[6] F. Davidoff, M. Dixon-Woods, L. Leviton, S. Michie, Demystifying theory and its use in improvement, BMJ Qual. Saf. 24 (3) (2015) 228–238, doi:10.1136/bmjqs-2014-003627.
[7] D.W. Eccles, G. Arsal, The think aloud method: What is it and how do I use it? Qual. Res. Sport Exerc. Health 9 (4) (2017) 514–531, doi:10.1080/2159676X.2017.1331501.
[8] J.P. Goetz, M.D. LeCompte, Ethnography and Qualitative Design in Education Research, Academic Press, 1984.
[9] K. Hankiewicz, Ergonomic context of the user interface of modern enterprise management systems, Eur. Res. Stud. J. 24 (5) (2021) 140–151, doi:10.35808/ersj/2708.
[10] M. Hertzum, Usability and user experience, in: Usability Testing: A Practitioner's Guide to Evaluating the User expErience, Springer International Publishing, 2020, pp. 9–14, doi:10.1007/978-3-031-02227-2_2.
[11] K. Hornbæk, Dogmas in the assessment of usability evaluation methods, Behav. Inf. Technol. 29 (1) (2010) 97–111, doi:10.1080/01449290801939400.
[12] E. Jaakkola, Designing conceptual articles: Four approaches, AMS Rev. 10 (1–2) (2020) 18–26, doi:10.1007/s13162-020-00161-0.
[13] K. Kager, J.P. Mynott, Applying a conceptual model to understand quality in Lesson Study: a cross-case usability test. (2024).
[14] H. Keenan, S. Duke, H. Wharrad, G. Doody, R. Patel, Usability: an introduction to and literature review of usability testing for educational resources in radiation oncology, Tech. Innov. Patient Support Radiat. Oncol. 24 (2022) 67–72, doi:10.1016/2Fj.tipsro.2022.09.001.
[15] S. Khan, R. VanWynsberghe, Cultivating the under-mined: cross-case analysis as knowledge mobilization, Forum Qual. Soc. Res. 9 (1) (2008) Article 4 http://www.qualitative-research.net/index.php/fqs/article/view/334/730.
[16] K. McChesney, J.M. Aldridge, What gets in the way? A new conceptual model for the trajectory from teacher professional development to impact, Prof. Dev. Educ. 47 (5) (2021) 834–852, doi:10.1080/19415257.2019.1667412.
[17] P. McMillen, D. Pehrsson, Improving a counselor education web site through usability testing: The Bibliotherapy Education Project, Couns. Educ. Superv. 49 (2) (2009) 122–136, doi:10.1002/j.1556-6978.2009.tb00092.x.
[18] S. Merriam, E. Tisdell, Qualitative Research: a Guide to Design and Implementation, 4th ed., John-Wiley & Sons, 2016.
[19] P. Morville, User experience design. (2004). Available at: https://semanticstudios.com/user_experience_design [Accessed 15th February 2024].
[20] J. Nielsen, Why you only need to test with 5 users. (2000). Available at: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/[Accessed 21st February 2024].
[21] J. Nielsen, Usability 101: Introduction to usability. Nielsen Norman. (2012). https://www.nngroup.com/articles/usability-101-introduction-to-usability/.
[22] M. Patton, Qualitative Evaluation and Research Methods, 2nd ed., Sage, 1990.
[23] W. Prussak, K. Hankiewicz, Quality in use evaluation of business websites: ergonomics in contemporary enterprise, in: Agile Enterprise: Concepts and Some Results of Research, IEA Press, 2007, pp. 84–91.
[24] W. Quesenbery, Balancing the 5Es: Usability. Clutter IT Journal, 17(2) (2004) 4-11. https://www.whitneyquesenbery.com/articles/5es-citj0204.pdf.

[25] D. Radovic, L. Black, J. Williams, C. Salas, Towards conceptual coherence in the research on mathematics learner identity a systematic review of the literature, Educ. Stud. Math. 99 (2) (2018) 1–22, doi:10.1007/s10649-018-9819-2.
[26] S. Richardson, R. Mishuris, A. O'Connell, D. Felstein, R. Hess, P. Smith, L. McCullagh, T. McGinn, D. Mann, Think aloud" and "Near Live" usability testing of two complex clinical decision support tools, Int. J. Med. Inform. 106 (2017) 1–8, doi:10.1016/j.ijmedinf.2017.06.003.
[27] S. Sridharan, A. Nakaima, Ten steps to making evaluation matter, Eval. Program Plann. 34 (2) (2011) 135–146, doi:10.1016/j.evalprogplan.2010.09.003.
[28] S.J. Tracy, Qualitative quality: Eight "big-tent" criteria for excellent qualitative research, Qual. Inq. 16 (10) (2010) 837–851, doi:10.1177/1077800410383121.
[29] R. Whittlemore, S. Chase, C. Mandle, Validity in qualitative research, Qual. Health Res. 11 (4) (2001) 522–537.
[30] R.K. Yin, Case Study Research: Design and Methods, Sage, 2014 2nd Ed., Vol. 5.