Student teachers' evaluative beliefs and quality criteria concerning a digital learning and test environment for the didactics of algebra

Meeri-Liisa Beste[1]; Bianca Wolff[1]; Boris Girnat[1]; Joaquin Veith[2]

1 University of Hildesheim, Germany
2 University of Leipzig, Germany

Abstract: The subject of this article is the evaluation of the test format that will be integrated into a digital self-learning platform in the future. On the basis of a qualitative preliminary study and taking into account motivation theories, a questionnaire was developed for the evaluation, which records the quality characteristics of the test formats as a self-assessment of the participants at four measurement points in the associated course. The data were analysed using a longitudinal study and exploratory cluster analysis. Surprisingly, there were no significant changes in the assessment of the quality of the test environment, even though test scores declined. Thus, the digital test environment seems to be perceived positively regardless of performance. The cluster analysis yields a solution with two clusters: a large cluster containing the participants with consistently high test scores and a small cluster in which the test scores (strongly) decrease during the four measurement points. It is noteworthy that the two clusters hardly differ in their positive assessment of the digital examination format.

Keywords: mathematics education, teacher students, digital learning environments, motivation, evaluation

Contact: boris.girnat@uni-hildesheim.de

1 Introduction

Digital learning environments have become widely used as a new teaching tool in recent years (Cleveland-Innes et al., 2024). Their use ranges from primary and secondary schools to universities and adult education. While research in mathematics education is dedicated to this new development and has already worked on a variety of different research topics, it has predominantly focused on digital learning environments used for mathematics teaching rather than the learning process, especially on digital learning environments used for mathematics teaching at school or university level (Cevikbas & Kaiser, 2021). Although the research on digital learning environments has grown, there are still research gaps in many areas, including evaluating learning environments (cf. Engelbrecht & Borba, 2024). This article aims to contribute to this topic by evaluating a course on the didactics of algebra within the context of teacher training at the university, using the participants' beliefs as the starting point for this evaluation.

Our study is connected to a long-term research and development project as part of revising the study regulations, in which traditional teacher training courses are to be supported by digital learning environments. The new study regulations add digital learning environments to six compulsory traditional courses. The first of these courses, with digital support, the Didactics of Algebra, was held in 2023. Therefore, this article refers to this course and presents the evaluation results of the digital learning environment used here. The fact that this course was "supported" by a digital learning environment means the following: The course consists of a traditional lecture and an online environment with two functions. Firstly, it provides materials for revision, in-depth study, and independent learning. It is updated weekly for each section of the lecture. Secondly, it replaces the traditional way of passing the exam: in the years before the online environment, the course was passed by writing four essays. Four online tests have now replaced these essays.

The central theme of our study is the participants' attitudes and beliefs towards the online environment, particularly their perception and acceptance of the online tests to pass the course. For this purpose, the 110 participants' evaluative beliefs were collected by a questionnaire with each test and examined for changes in a longitudinal study. These beliefs and the test results were linked to background variables available about the participants from a previous semester.

The aim was, on the one hand, to evaluate the new online offer, including its acceptance and, on the other hand, to classify and explain the participants' beliefs in the light of various background variables. The purpose of the second goal is not only to focus on a singular learning environment but also to examine connections that may be generalisable beyond this one case. Beliefs and some appropriately selected background variables are considered suitable for doing this, as they are independent of the specific learning environment.

2 The online environment for the didactics of algebra

Since 2022, digital learning environments have been set up at the mathematical institute at the University of Hildesheim to support six curricular courses in the teacher training programme (University of Hildesheim, 2022), namely for algebra, geometry, and analysis as well as for the corresponding didactic courses on the didactics of algebra, geometry, and functional thinking (Beste et al., 2023). This project follows a pilot study in 2019 that established a digital learning environment for the beginners' lecture "Introduction to the Didactics of Mathematics" (Kober & Girnat, 2021). The six learning environments of the current project are aimed at further courses from the second semester onwards. Student teachers attend them with a focus on primary or lower secondary education.

The learning platform for the didactics of algebra has been set up as the first of the six planned environments and was offered for the first time in the summer semester of 2023 for the primary and secondary schools teaching degree programs. The didactics of algebra was offered for the first time as a new course in teacher training at the University of Hildesheim as part of the change to the study regulations. The establishment of such a course is relevant for student teachers for the following reasons: In their bibliometric analysis, Veith et al. (2022) have shown that a great deal of research has been carried out in algebra education over the last 20 years. A performance analysis shows "the vastly growing output of research in algebra education" (p. 13). It is, therefore, timely to offer a course on the didactics of algebra as the field of research continues to grow. Secondly, algebraic thinking is central to understanding mathematics. It involves recognising patterns and understanding the relationships between numbers, objects and shapes (Windsor, 2009). However, algebra is often challenging for learners and is associated with some misconceptions. Much of the time is spent on so-called 'letter arithmetic' (Malle, 1993). In order to promote algebraic thinking in students and to be able to deal with misconceptions, student teachers should develop strategies in the didactics of algebra that specifically address these aspects (Windsor, 2010).

As part of the project, an online learning environment was developed to provide optimal support for this lecture. As both the online learning environment and the lecture were being used for the first time, an evaluation was of particular interest to improve the quality of teaching.

This is why this course was chosen for this work. The learning environment accompanies the corresponding lecture and includes topics such as early algebra, patterns and structures, arithmetic operations, the introduction of placeholders and variables, terms and algebraic expressions, linear and quadratic equations, functional thinking, and algebra in applied contexts. The learning content is represented using different media: videos, PDF files, articles, and external websites.

The learning environment consists of four online tests, comprising a total of 94 tasks. Each task must be completed within two weeks at various times throughout the semester to pass the lecture. The tasks of the four tests were developed based on the lecture content in accordance with Anderson and Krathwohl's learning taxonomy. This automatically analysable task covers the cognitive levels of remembering, understanding, and applying, with a particular focus on application-related tasks (Anderson & Krathwohl, 2001). The question types used are multiple-choice, cloze, and matching. To evaluate the test tasks, an online questionnaire was provided for all four tests. The online questionnaire consisted of the same items at all four measurement times.

The establishment and operation of the learning platform for the didactics of algebra build on the research of Lagrange and Kynigos (2014). The platform's structure reflects the endeavour to incorporate technology in a contextualised way to enable students to develop a deeper understanding of algebraic concepts while improving their learning experience through the constant feedback between the test results and the questionnaire (Lagrange & Kynigos, 2014).

3 Theoretical background and items of the questionnaire

The evaluation of the online tasks is based on two sources: The pilot project related to the beginners' lecture "Introduction to the Didactics of Mathematics" in 2019 was already based on online tests but was not supported by a digital learning environment complementing the lecture. These tests were evaluated back in 2019. Since there was no experience evaluating online tests, the participants were asked only two questions with free text answers: "What did you like about the online tests?" and "What do you think is negative about the online tests?". Three statements were most frequently mentioned as positive features: The online tests cover the content of the lectures more broadly than essays, exams, or other common ways to pass the exam; the online tests allow independent, flexible processing over the course of the semester; the online tests are welcome alternative forms of assessment instead of the usual written work or exams that dominate studies. The three most frequently mentioned negative statements were: The test items were too difficult; completing the tasks was too time-consuming; and some questions were worded ambiguously (Kober & Girnat, 2021). These statements were the initial point for the further quantitative evaluation and were taken into account to create closed-ended items for evaluating the algebra didactics course in 2023.

The second source of the questionnaire is based on theoretical considerations used in other areas of belief research, especially those from motivation research (Hannula et al., 2016). Our approach to motivation is based on the works of Deci and Ryan (1985; 2008) and Ryan and Deci (2000).

Deci and Ryan distinguish several types of motivation, particularly intrinsic, extrinsic, and instrumental. Intrinsic motivation is defined as the doing of an activity for its inherent satisfaction rather than for some separable consequences; extrinsic motivation is a motivation that is driven to gain external rewards or to avoid punishments; and instrumental motivation is linked to a practical or pragmatic reason, i.e. it refers to the motivation to engage in an activity because it is personally important or relevant to one's goals, values, or identity. For our purposes, extrinsic motivation does not seem to be of interest, as the evaluation of tasks should not lie in rewards or punishments in the social context of these tasks but in the tasks and the characteristics, identity, and goals of the test participants. For this reason, we only included intrinsic and instrumental motivation when creating the closed items of our questionnaire. In intrinsic motivation, by definition, the focus is on interest in the subject. For instrumental motivation, we decided to set goals related to the current studies and the future career as teachers.

The connection of current and future (professional) goals also takes up an idea from Klafki's pedagogical theory, the so-called "learning goal analysis": Here, too, the value of newly learned knowledge is measured by its current and future significance for the learner's life (Klafki, 1995, 1996).

The items listed in Table 1 were created according to these considerations: Items Q1 to Q3 take up the most frequently mentioned positive answers from the evaluation of the pilot project in 2019. Item Q4 addresses intrinsic motivation, while items Q5 and Q7 address instrumental motivation and, at the same time, future significance in Klafki's sense. When formulating items Q5 and Q7, we particularly took into account a research result from Simons et al. (2004) that instrumental motivation can be better captured the more concretely the desired goal - in this case the teaching profession - is formulated. Item Q6 addresses a learning goal that is often cited as an advantage of digital learning environments, namely the opportunity to learn in a self-directed manner according to individual prior knowledge (Attert & Holmes, 2022; Armborst-Weihs et al., 2018; Leutner, 2022).

The Comp item compares the new exam format with traditional alternatives. The Diff and Time items address the negative responses collected in the pilot study evaluation: Are the test tasks too difficult? Do the tests take too much time?

Table 1. Items used in the questionnaire with response options and coding

Item	**Item**	**Response options and their coding**
Q1	I learned a lot by completing the tests.	... strongly disagree (0) ... tend to disagree (1) ... tend to agree (2) ... completely agree (3)
Q2	The test tasks broadly cover the content of the lecture.	
Q3	To complete the test, I intensively studied the content of the lecture.	
Q4	The test tasks are interesting in terms of content.	
Q5	The topics of the test tasks are relevant for my future career as a teacher.	
Q6	By completing the test, I was encouraged to deal with the learning content independently.	
Q7	The topics of the test tasks are important for my further studies.	
Comp	Online tests are a better way to earn coursework than other typical requirements (such as essays, assignments, exams, portfolios, etc.).	
Diff	The test tasks are ...	... much too easy (0) ... rather too easy (1) ... appropriate (2) ... rather too difficult (3) ... much too difficult (4)
Time	I need approximately the following time per week to repeat the lecture content:	... 0 minutes (0) ... 0 to 30 minutes (1) ... 30 minutes to an hour (2) ... one to two hours (3) ... more than two hours (4)

4 Research questions

The aim of the study is to use the items as measuring instruments for the acceptance of the online tests. The question is whether the positively worded items Q1 to Q7 form a common scale. This is supported by the fact that each of them captures the quality characteristics of the environment; the argument against this is that they come from different background theories: intrinsic motivation, instrumental motivation, and exploratory statements from the pilot study. It will be further investigated once it is clear whether a scale can be formed from the items. The focus is on possible changes between the four measuring times and connections to the background variables of the participants (such as previous test results). More precisely, the research questions can be formulated as follows:

- RQ1: How are the items of the questionnaire correlated?
- RQ2: Can the items be used to form a common scale that can be used to measure the "overall acceptance" of the digital learning environment?
- RQ3: Do the measured values (score) of the items or an overall scale change over the four measurement points?
- RQ4: Are there gender differences or other types of group differences (primary education vs. secondary education)?
- RQ5: How do the items or an overall scale correlate with the background variables of the participants (such as previous achievements in mathematics and mathematics didactics)?
- RQ6: Can explorative cluster analysis be used to find groups that are homogeneous in themselves but differ from each other?

5 Sample and methods

110 students attended the lecture on the digital learning environment in the summer semester of 2023. 78 of them indicated their gender as "female" and 32 as "male." 88 participants stated that the focus of their study was "primary school" (grades 1 to 4), and 22 indicated that the focus was "secondary school" (grades 5 to 10).

Over the course of the semester, the students completed four online tests with the developed tasks. After each of these tests, a digital questionnaire survey with the items presented in Table 1 is available to the students voluntarily. This method allows for flexible and location-independent participation (Dillman et al., 2014; Evans & Mathur, 2018). The data collected provides valuable insights into motivation and test scores, which allows targeted adjustments to the learning environment. Specific aspects of the learning environment can be improved by making targeted adjustments based on the feedback, which continuously adapts the learning experience to the learners' needs.

As the course was offered for the first time, there were no repeat participants; all participants were in their second semester. The participants had attended two previous courses, the results of which are of interest as background variables for the evaluation of the questionnaire: a lecture "Introduction to Mathematics" (M1) and a lecture "Introduction to the Didactics of Mathematics" (D1).

The programme R was used to analyse the data (R core team, 2024). In addition to the standard methods of R, the *lme4* package (Bates et al., 2015) was used to examine changes between the four measurement points using linear multilevel models (Finch & Bolin, 2019). The explorative cluster analysis was done using the R package *mclust* (Scrucca et al., 2023). Missing data (less than 3% per item) were handled by multiple imputations (van Buuren, 2018) using the package *mice* (van Buuren & Groothuis-Oudshoorn, 2011).

6 Results

Firstly, it is analysed how the individual items of the questionnaire correlate with each other (RQ1). An exploratory factor analysis has shown that the three items of the online tests (*Comp*), the difficulty (*Diff*) and the time (*Time*) should be treated as single items and that all other items can be combined into a common scale, which is considered as a scale for the general quality of the offer (RQ2) (Cronbach's alpha t0: 0.78; t1: 0.76; t2: 0.73; t3: 0.84; Cronbach, 1951). This scale is labelled as *Sc.qual*.

Table 2 shows the development of the results over time. The mean values are given for the questionnaire items and the "quality scale". In addition, the percent test score for the tests is given (Cronbach's alpha of the tests: t0: 0.86; t1: 0.88; t2: 0.92; t3: 0.92). Significance is indicated by the common asterisks (* for $p<0.05$, ** for $p<0.01$, and *** for $p<0.001$).

Table 2. Means related to the four measuring times

Items/Scale	**Measuring time (Means)**				**Parameters**	
	t0	t1	t2	t3	R^2	Theoretical mean
Test	0.70	0.66	0.59	0.59	0.113***	0.50
Q1	1.80	1.83	1.92	1.76	0.007	1.50
Q2	2.13	2.09	2.04	2.11	0.002	1.50
Q3	2.20	2.31	2.19	2.04	0.016	1.50
Q4	1.77	1.79	2.07	1.78	0.026	1.50
Q5	2.11	1.96	2.01	1.87	0.013	1.50
Q6	2.16	2.17	2.35	2.27	0.010	1.50
Q7	1.96	1.95	2.08	1.80	0.023	1.50
Sc.qual	2.00	2.02	2.10	1.94	0.013	1.50
Comp	2.65	2.72	2.65	1.53	0.002	1.50
Diff	1.29	1.65	1.57	1.53	0.062***	2.00
Time	2.61	2.56	2.59	2.49	0.003	2.00

Regarding Table 2, it is noteworthy that the means of all items (except for difficulty) are above the theoretical mean (of 1.50 and 2.00, respectively) at all measurement times (RQ3). This supports the interpretation that the online tests are generally perceived as positive, which is confirmed by the high mean values of the “online tests item” (*Comp*). It is also worth noting that only the test scores and the subjective assessment of difficulty (*Diff*) changed significantly over time, i.e., the rating of the online tests remains practically constant (despite the decreasing test scores and increasing difficulty rating). This suggests that online tests are viewed positively regardless of performance development.

Another research question concerns group differences (RQ4). The sample includes the two group variables gender and focus of study (primary vs. secondary). No significant group differences were found for these two grouping variables, and there were no significant differences in the temporal development of these grouping variables over time at the four measurement points.

Correlations are next examined (RQ5). As items Q1 to Q7 have been combined to form the scale *Sc.qual*, these items are not considered individually but only the scale. The remaining three items *Comp*, *Diff* and *Time* are included in the evaluation as individual items. In addition, the results of two other lectures are included as background variables, namely the results from the lecture “Introduction to Mathematics” (M1) and from the lecture “Introduction to the Didactics of Mathematics” (D1). The results of these two lectures, like the results of the online tests, are available as percentage values. In order to keep the number of correlations manageable, the average values are formed from all four measurement points.

Table 3. Correlations between the scale, the test score, the single items, and the backgrounds variables (coefficients in the upper triangle matrix; significance levels in the lower half)

	Test	**Sc.qual**	**Comp**	**Diff**	**Time**	**M1**	**D1**
Test	1	0.19	0.03	0.03	-0.11	0.23*	0.77***
Sc.qual	0.051	1	0.29**	0.17	0.21*	-0.03	0.19
Comp	0.741	0.002	1	0.00	0.18	-0.01	0.07
Diff	0.738	0.090	0.998	1	-0.25*	-0.06	-0.04
Time	0.251	0.032	0.068	0.012	1	-0.20	-0.05
M1	0.041	0.812	0.902	0.605	0.072	1	0.29***
D1	<0.001	0.052	0.507	0.698	0.621	<0.001	1

Both the significant and the non-significant correlations in Table 3 are interesting: the *Sc.qual* scale correlates slightly positively, but (almost) not significantly, with the results of the online tests (*Test*), i.e. even participants with lower test results do not necessarily have a negative correlation with the online test evaluation. The correlation between *Comp* and *Test*, which is practically zero, can be interpreted in a similar way: the preferences for online tests over other exam formats does not depend on the results in the tests. Overall, online tests are an accepted format for participants, regardless of their performance. The moderate to high correlations between the online tests and previous mathematical and didactic performance (M1 and D1) are not surprising.

Table 4. Means related to the two clusters

Items/Scale	**Cluster**	**Measuring time (Means)**			
		t0	**t1**	**t2**	**t3**
Test	1 2	0.70 0.44	0.67 0.23	0.60 0.10	0.60 0.05
Sc.qual	1 2	2.02 1.86	2.01 1.93	2.04 1.85	1.93 1.90
Comp	1 2	2.67 2.53	2.73 2.73	2.60 2.66	2.73 2.59
Diff	1 2	1.53 1.28	1.47 1.63	0.93 1.49	1.00 1.45
Time	1 2	2.62 2.60	2.47 2.93	2.51 2.33	2.47 3.27
M1	1 2	0.58 0.37	-	-	-
D1	1 2	0.70 0.26	-	-	-

The final section of our evaluation relates to an exploratory cluster analysis (RQ6). The *mclust* package provides automatic model selection and proposes a solution with two clusters. The first cluster is very large with 95 members (86.4%), the second is very small with 15 members (13.6%). Table 4 lists the characteristics of these clusters, i.e. the means across the test scores, the scale Sc.qual, the items Comp, Diff, and Time and the covariates M1 and D1.

Regarding Table 4, the two clusters can be interpreted in an obvious way: Cluster 1 includes the participants who maintain a high level in the online tests. Cluster 2 includes the participants who start with low test results and drop to close to zero by the fourth measurement point. The large difference in performance in the covariates (in mathematics M1 and in didactics D1) shows that the previous performance indicates whether or not the course on didactics of algebra course will be successfully completed in the second semester. In contrast, it is rather surprising that both groups rate the format of the online tests (*Comp*) positively and without major differences and that the quality of the test tasks (*Sc.qual*) is also evaluated similarly by both groups.

6 Conclusions

The main objective of the study was to find a way to evaluate a digital learning environment. The items Q1 to Q7 are so strongly correlated (RQ1) that it is possible to form a common scale with good reliability (with Cronbach's alpha of 0.73 to 0.84, depending on the time of measurement). This scale evaluates the "overall acceptance" of the environment (RQ2). This result was not necessarily expected, as the items were partly developed exploratively from a pilot study, i.e. without any theoretical background and with different theoretical backgrounds: with reference to intrinsic and instrumental motivation (Deci & Ryan, 1985; Ryan & Deci, 2000) and with the inclusion of Klafki's learning goal theory (1995, 1996). This result can be seen as positive, as this scale covers a wide range of quality characteristics with items from various theoretical and content-related aspects. Surprisingly, the mean values of the scale and the individual items (except for the assessment of difficulty) do not change significantly over all four measurement points in the semester (RQ3), i.e. the assessment of the learning environment remains stable, even if the test results and the assessment of the difficulty change. The positive aspect is that the learning environment appears to be valued independently of performance (as shown by the high values on all items).

No significant group differences (RQ4) indicate a general acceptance of the examination format. This general acceptance is echoed in previous research, which indicates that digital environments are well-received alternatives to traditional assessments (Kober & Girnat, 2021). Significant correlations exist between the previously achieved performance in mathematics and didactics (M1 and D1) and current test results. This suggests that the knowledge and skills previously acquired by the participants influence their current results (RQ5).

Interestingly, the exploratory cluster analysis (RQ6) leads to an easy-to-interpret solution with two clusters: The much larger cluster (86.4%) includes the participants whose test level is high throughout the semester. Cluster 2 (13.6%) includes the participants who start with low test scores and drop to close to zero by the fourth measurement point. It should be noted that the second cluster can already be identified from the preliminary work, i.e. from the courses "Introduction to Mathematics" and "Introduction to Didactics of Mathematics" (Kober & Girnat, 2021). Surprisingly, the members of both clusters rate the examination offer similarly and positively despite the differences in performance.

The study is only the first step in evaluating the whole project: out of six lectures supported by a digital learning environment, only the first has been evaluated. In addition, the evaluation only refers to the test format, not the entire learning environment. It should also be noted that the study used a relatively small sample of 110 participants over a single semester, suggesting that future research should gather broader feedback in more diverse contexts to earn experiences through contextualized technology integration, as noted by Lagrange and Kynigos (2014). The question is whether the results will remain stable or reproducible in larger samples and over a longer period of time. This is particularly the case for the cluster solution. In terms of content, it is well understandable that there could be a cluster with declining performance each year. Still, it is questionable whether this cluster behaves similarly regarding (positive) evaluation of the digital learning environment and in terms of (low) prior performance.

Research ethics

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by M.-L. B., B.W., J. V. and B. G. All authors have read and agreed to the published version of the manuscript.

Artificial intelligence

Artificial intelligence (DeepL) was used to revise the language of the article.

Funding

No funds, grants, or other support was received.

Institutional review board statement

All participants in the study were of legal age and consented to the study on a voluntary basis.

Informed consent statement

Informed consent was obtained from all research participants

Data availability statement

In order to protect the privacy of the participants and the confidentiality of the information, we have decided not to make the data sets publicly accessible. However, we recognise the importance of transparency and traceability in scientific research and are therefore prepared to make the data available on reasoned scientific request. We ask that all requests be sent in writing to the correspondence address of the corresponding author, which is given in the article.

Conflicts of interest

The authors declare no conflicts of interest.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). Longman.

Armborst-Weihs, K., Böckelmann, C., & Halbeis, W. (2018). *Selbstbestimmt lernen – Selbstlernarrangements gestalten. Innovationen für Studiengänge und Lehrveranstaltungen mit kostbarer Präsenzzeit* [Self-determined learning – designing self-learning arrangements. Innovations for degree programmes and teaching events with valuable attendance time]. Waxmann. https://doi.org/10.5281/zenodo.1209167

Attard, C., & Holmes, K. (2022). An exploration of teacher and student perceptions of blended learning in four secondary mathematics classrooms. *Mathematics Education Research Journal, 34*, 719–740. https://doi.org/10.1007/s13394-020-00359-2

Barana, A., & Marchisio, M. (2020). An interactive learning environment to empower engagement in mathematics. *Interaction Design and Architecture(s) Journal, 45*, 302–321. https://doi.org/10.55612/s-5002-045-014

Barana, A., Conte, A., Fissore, C., Marchisio, M., & Rabellino, S. (2019). Learning analytics to improve formative assessment strategies. *Journal of e-Learning and Knowledge Society, 15*, 75–88. https://doi.org/10.20368/1971-8829/1135057

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beste, M.-L., Wolff, B., & Veith, J. (2023). Entwicklung einer Selbstlern-Plattform im Projekt „Digital C@mpus-le@rning“ der Universität Hildesheim [Development of a self-learning platform in the "Digital C@mpus-le@rning" project at the University of Hildesheim]. *GDM-Mitteilungen, 114*. Retrieved April 15, 2025, from https://ojs.didaktik-der-mathematik.de/index.php/mgdm/article/view/1128/1341

Cevikbas, M., & Kaiser, G. (2021). A systematic review on task design in dynamic and interactive mathematics learning environments (DIMLEs). *Mathematics, 9*(4). https://doi.org/10.3390/math9040399

Cleveland-Innes, M. F., Stenbom, S., & Garrison, D. R. (Eds.). (2024). *The design of digital learning environments: Online and blended applications of the community of inquiry* (1st ed.). Routledge. https://doi.org/10.4324/9781003246206

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press. https://doi.org/10.1007/978-1-4899-2271-7

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne, 49*(3), 182–185. https://doi.org/10.1037/a0012801

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). John Wiley & Sons.

Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. *ZDM Mathematics Education, 56*, 281–292. https://doi.org/10.1007/s11858-023-01530-2

Evans, J. R., & Mathur, A. (2018). The value of online surveys: a look back and a look ahead. *Internet Research,* 28(4), 854-887. https://doi.org/10.1108/IntR-03-2018-0089

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel modeling using R* (2nd ed.). CRC Press. https://doi.org/10.1201/9781351062268

Hannula, M., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanim, E., Goldin, G. (2016). *Attitudes, beliefs, motivation and identity in mathematics education*. Springer. https://doi.org/10.1007/978-3-319-32811-9_1

Hannula, M. S., Leder, G. C., Morselli, F., Vollstedt, M., & Zhang, Q. (2019). *Affect and mathematics education: Fresh perspectives on motivation, engagement, and identity*. Springer Nature. https://doi.org/10.1007/978-3-030-13761-8_1

Klafki, W. (1995). Didaktische Analyse als Kern der Unterrichtsvorbereitung [Didactic analysis as the core of preparation of instruction]. *Journal of Curriculum Studies, 27*(1), 13–30. https://doi.org/10.1080/0022027950270103

Klafki, W. (1996). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* [New studies on educational theory and didactics. Contemporary general education and critical-constructive didactics*]* (5th ed.). Beltz.

Kober, T., & Girnat, B. (2021). Eine digitale Lern- und Prüfungsumgebung zur Einführung in die Didaktik der Mathematik [A digital learning and examination environment for the introduction to the didactics of mathematics]. In B. Girnat (Ed.), *Mathematik lernen mit digitalen Medien und forschungsbezogenen Lernumgebungen – Innovationen in Schule und Hochschule* (pp. 71–95). Springer. https://doi.org/10.1007/978-3-658-32368-4_4

Lagrange, J.-B., & Kynigos, C. (2014). Digital technologies to teach and learn mathematics: Context and re-contextualization. *Educational Studies in Mathematics, 85*(3), 381–403. https://doi.org/10.1007/s10649-013-9525-z

Leutner, D. (2002). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme [Adaptivity and adaptability of multimedia teaching and information systems]. In L. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis* (pp. 114–125). Beltz.

Leder, G., Pehkonen, E., & Törner, G. (Eds.). (2002). *Beliefs: A hidden variable in mathematics education?* Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47958-3

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models—An introduction to factor, path, and structural equation analysis* (5th ed.). Routledge. https://doi.org/10.4324/9781315643199

Malle, G. (1993). Schülerfehler beim Umformen [Student error when transforming]. In E. C. Wittmann (Ed.), *Didaktische Probleme der elementaren Algebra* (pp. 160–187). Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-89561-5_7

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Information Age.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459–470. https://doi.org/10.1016/S0883-0355(99)00015-4

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

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

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.102

Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. Chapman & Hall/CRC. https://doi.org/10.1201/9781003277965

Simons, J., Vansteenkiste, M., Lens, W., & Lacante, M. (2004). Placing motivation and future time perspective theory in a temporal perspective. *Educational Psychology Review, 16*(2), 121–139. https://doi.org/10.1023/B:EDPR.0000026609.94841.2f

Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics learning and teaching* (pp. 127–146). Macmillan.

University of Hildesheim. (2022). *Neufassung der Studienordnung für das Fach Mathematik: Polyvalente Zwei-Fächer-Bachelor-Studiengänge* [New version of the study regulations for the subject mathematics: Polyvalent two-subject bachelor's degree programmes]. Retrieved April 20, 2025, from https://www.uni-hildesheim.de/qm/processmanagement/download.php?fileID=4996

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman & Hall/CRC. https://doi.org/10.1201/9780429492259

Veith, J. M., Beste, M.-L., Kindervater, M., Krause, M., Straulino, M., Greinert, F., & Bitzenbauer, P. (2023). Mathematics education research on algebra over the last two decades: Quo vadis? *Frontiers in Education, 8*, Article 1211920. https://doi.org/10.3389/feduc.2023.1211920

Windsor, W. (2009). Algebraic thinking—More to do with why, than x and y. In *Proceedings of the 10th International Conference "Models in Developing Mathematics Education"* (pp. 592–595). The Mathematics Education into the 21st Century Project.

Windsor, W. (2010). Algebraic thinking: A problem-solving approach. In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia* (pp. 665–672). MERGA.