

Mobile Learning in Medical Education: A Realist Evaluation of Usage, Context, and Exam Performance in a Curricular Setting

Jan Peter Herbert Holzhäuer, Joy Backhaus, Thiên-Trí Lâm, Sarah König

Submitted to: JMIR Medical Education
on: October 15, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	30
Figures	31
Figure 1.....	32
Figure 2.....	33
Multimedia Appendixes	34
Multimedia Appendix 1.....	35
Multimedia Appendix 2.....	35
Multimedia Appendix 3.....	35
Multimedia Appendix 4.....	35
Multimedia Appendix 5.....	35
Multimedia Appendix 6.....	35

Mobile Learning in Medical Education: A Realist Evaluation of Usage, Context, and Exam Performance in a Curricular Setting

Jan Peter Herbert Holzhäuer¹; Joy Backhaus¹ MSc; Thiên-Trí Lâm² MME, MD; Sarah König¹ MME, MD

¹ Institute of Medical Teaching and Medical Education Research Universitätsklinikum Würzburg Würzburg DE

² Institute for Hygiene and Microbiology University of Würzburg Würzburg DE

Corresponding Author:

Sarah König MME, MD

Institute of Medical Teaching and Medical Education Research
Universitätsklinikum Würzburg
Josef-Schneider-Str. 2/D6
Würzburg
DE

Abstract

Background: Mobile learning (mLearning) is widely used in higher education but research remains fragmented and context-insensitive. In undergraduate medical education, in which knowledge acquisition must align with competency-based training, mLearning may bridge formal and informal learning. Previous research has focused on two main areas: technology acceptance or intervention effectiveness, but rarely their integration. Using realist evaluation (RE), this study investigated the conditions under which mLearning supports outcomes in authentic curricula settings.

Objective: This study aimed to (1) assess the impact of a voluntary mLearning intervention on exam performance and (2) identify learner profiles and context–mechanism–outcome (CMO) patterns explaining differential engagement and outcomes.

Methods: A quasi-experimental study was conducted among fifth-semester undergraduate medical students at a German medical school across two consecutive summer semesters (2023 and 2024). A voluntary, app-based mLearning course in “Microbiology, Hygiene, and Virology” was delivered via the eSquirrel platform, featuring interactive tasks, gamification, and spaced-repetition features. Data sources included non-reactive app usage logs, baseline academic performance, demographics, and post-semester survey responses. Cluster analysis identified usage profiles, and CMO patterns were derived by triangulating usage, performance, and survey data. Neither randomization nor blinding was applied, to capture real-world curricular use.

Results: Of 245 eligible students, 220 (89.8%) participated; 110 (50%) used the app. In 2024 app users (n=64) outperformed non-users (n=46) in the oral microbiology exam (mean grade 2.3 vs 2.8; $t_{(100)}=1.90$, one-sided $P=.031$). After adjustment, these differences were largely explained by baseline academic performance, with limited evidence for an independent app effect. Cluster analysis of app users identified three engagement profiles: (1) continuous low-intensity users (n=60), (2) intensive, exam-oriented users (n=31), and (3) early, disengaging users (n=19). Cluster 2 reported the highest satisfaction, perceived learning gains, and exam performance. A refined program theory suggested that when learners with sufficient baseline ability and strategic study habits engage “just-in-time” (Context), mechanisms such as motivation, spaced repetitions, microlearning, and gamified feedback (Mechanism) lead to improved exam performance (Outcome).

Conclusions: The effectiveness of mLearning is not universal but varies with learner context and engagement strategy. Strategic, exam-oriented use—driven by motivation, enjoyment, and self-regulation—yielded the strongest benefits. Passive usage data provided valuable insights into authentic learning behavior, while quasi-experimental designs require contextual adjustment to clarify causal relationships. Effective mLearning interventions should be both educationally useful and engaging, fostering enjoyment alongside learning outcomes. Future research should address equity concerns, as higher-performing students benefited most, and explore adaptive approaches to support diverse learners.

(JMIR Preprints 15/10/2025:85892)

DOI: <https://doi.org/10.2196/preprints.85892>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>

Original Manuscript

Manuscript title

Mobile Learning in Medical Education: A Realist Evaluation of Usage, Context, and Exam Performance in a Curricular Setting

Paper type

Original Paper

Authors' information/metadata

Jan P. H. Holzhäuer:

University Hospital Würzburg, Institute of Medical Teaching and Medical Education Research (Würzburg, Germany)

Joy Backhaus:

University Hospital Würzburg, Institute of Medical Teaching and Medical Education Research (Würzburg, Germany)

Thiên-Trí Lâm:

Julius Maximilian University Würzburg, Institute of Hygiene and Microbiology (Würzburg, Germany)

Sarah König:

University Hospital Würzburg, Institute of Medical Teaching and Medical Education Research (Würzburg, Germany)

Corresponding author

Sarah König

Josef-Schneider-Str. 2/D6

97080 Würzburg (Germany)

Mailto: koenig_s7@ukw.de

Fon: +49 931 201 55210

Fax: +49 931 201 55222

Length of paper

Titel: 122 characters (including spaces)

Abstract: 423 words (including headlines)

Text: 5122 words

Mobile Learning in Medical Education: A Realist Evaluation of Usage, Context, and Exam Performance in a Curricular Setting

Abstract

Background

Mobile learning (mLearning) is widely used in higher education but research remains fragmented and context-insensitive. In undergraduate medical education, in which knowledge acquisition must align with competency-based training, mLearning may bridge formal and informal learning. Previous research has focused on two main areas: technology acceptance or intervention effectiveness, but rarely their integration. Using realist evaluation (RE), this study investigated the conditions under which mLearning supports outcomes in authentic curricula settings.

Objective

This study aimed to (1) assess the impact of a voluntary mLearning intervention on exam performance and (2) identify learner profiles and context–mechanism–outcome (CMO) patterns explaining differential engagement and outcomes.

Methods

A quasi-experimental study was conducted among fifth-semester undergraduate medical students at a German medical school across two consecutive summer semesters (2023 and 2024). A voluntary, app-based mLearning course in “*Microbiology, Hygiene, and Virology*” was delivered via the *eSquirrel* platform, featuring interactive tasks, gamification, and spaced-repetition features. Data sources included non-reactive app usage logs, baseline academic performance, demographics, and post-semester survey responses. Cluster analysis identified usage profiles, and CMO patterns were derived by triangulating usage, performance, and survey data. Neither randomization nor blinding was applied, to capture real-world curricular use.

Results

Of 245 eligible students, 220 (89.8%) participated; 110 (50%) used the app. In 2024 app users ($n=64$) outperformed non-users ($n=46$) in the oral microbiology exam (mean grade 2.3 vs 2.8; $t_{63,0}=1.90$, one-sided $P=.031$). After adjustment, these differences were largely explained by baseline academic performance, with limited evidence for an independent app effect. Cluster analysis of app users identified three engagement profiles: (1) continuous low-intensity users ($n=60$), (2) intensive, exam-oriented users ($n=31$), and (3) early, disengaging users ($n=19$). Cluster 2 reported the highest satisfaction, perceived learning gains, and exam performance. A refined program theory suggested that when learners with sufficient baseline ability and strategic study habits engage “just-in-time” (*Context*), mechanisms such as motivation, spaced repetitions, microlearning, and gamified feedback (*Mechanism*) lead to improved exam performance (*Outcome*).

Conclusion

The effectiveness of mLearning is not universal but varies with learner context and engagement strategy. Strategic, exam-oriented use—driven by motivation, enjoyment, and self-regulation—yielded the strongest benefits. Passive usage data provided valuable insights into authentic learning behavior, while quasi-experimental designs require contextual adjustment to clarify causal relationships. Effective mLearning interventions should be both educationally useful and engaging, fostering enjoyment alongside learning outcomes. Future research should address equity concerns, as higher-performing students benefited most, and explore adaptive approaches to support diverse learners.

Keywords

Medical Education; Mobile Learning; Realist Evaluation; Usage Patterns; Digital Learning Tools,

Technology Acceptance; Student Engagement; Educational Technology; Context Factors



Introduction

Mobile learning (mLearning) has become a prominent but also debated concept in technology-enhanced education. Research in this area continues to face enduring conceptual and methodological challenges.

First, definitions of mLearning remain inconsistent and have shifted alongside changing technological and pedagogical paradigms[1–6]. In the early 2000s, mLearning was mostly defined as “learning anytime and anywhere using mobile devices”[7–12]. This device-oriented view initially aimed to distinguish mobile from electronic learning (eLearning) and other stationary formats. With the widespread adoption of smartphones, early visions of an all-in-one mLearning device—always connected, interactive, and multimedia-capable—appeared to become reality[13,14]. More recent definitions are increasingly theory-based approaches and adopt a learner-centered perspective. They emphasize context-awareness, personalization, and continuity between physical and digital learning spaces[4,13–24]. In this view, mLearning is no longer just a technological mode of delivery but reflects the educational dynamics of a mobile society[25–27]. Mobile learning represents the typical learning mode of *Homo mobilis*, the mobile individual, within today’s mobile society[20,28–31]. Notably, some scholars have further argued that mLearning may also encompass traditional learning activities—such as reading a textbook while commuting[28,32–34]. Secondly, the rapid pace of technological innovation quickly renders many study results outdated, limiting their generalizability[35]. What was once counted mLearning—for example receiving an SMS from a teacher[14]—differs markedly from today’s artificial intelligence (AI)-powered mobile apps or augmented reality tools. Third, the contexts in which mLearning has been evaluated are highly diverse[14,15,36]: Existing studies have made valuable contributions by examining a wide range of interventions across educational levels, settings, and technologies. Yet, the very diversity—in intervention types, instructional content, outcome measures, and study populations—makes it difficult to draw broad conclusions.

Although much of the literature is optimistic[37–42], mLearning has not delivered the anticipated educational transformation[43]. Thus, more context-sensitive research is needed to understand how and under what conditions mLearning can be effectively integrated into higher education[44]. This challenge is particularly pressing in undergraduate medical education[45], in which curricula are shifting toward a hybrid and individualized format and in a broader move from knowledge transmission to competency-based learning[46,47]. As students are expected to develop professional competencies rather than merely acquire facts, a didactic gap emerges: foundational theoretical knowledge must be acquired through self-directed learning in informal contexts. mLearning may help bridging this didactic gap by supporting learners in connecting theoretical knowledge with formal competency development[48–50]. By providing content access “anytime and anywhere,” it can scaffold informal learning episodes that complement clinical training and enhance readiness for professional practice. Consistently, informal learning spaces are gaining importance alongside structured, institutional instruction[51], and mLearning can contribute to more integrated, context-sensitive learning trajectories in medical education[52].

Within mLearning research, two major lines of inquiry have received particular attention[53]: (1) studies on the overall effectiveness of mLearning interventions and (2) investigations into psychological and contextual factors shaping learners’ adoption. These perspectives have rarely been integrated under authentic curricular conditions. As a result, there is little understanding of how learners actually engage with mobile tools, how contextual factors shape this engagement, and what the outcomes are ultimately.

To address this gap, we applied the realist evaluation (RE) framework[54] to examine the interplay between learner context, engagement patterns, and academic outcomes. Using the context–mechanism–outcome (CMO) model[55,56], we conceptualize mLearning as a contingent process in

which learner-specific contexts (C) activate engagement mechanisms (M) that, in turn, have measurable outcomes (O).

Our study pursued two main objectives: (1) To test whether voluntary participation in a mLearning course improved exam performance within routine curricular delivery, and (2) to explore how learner contexts influenced engagement patterns by deriving usage profiles from in-app behavior and linking them to contextual variables and academic outcomes.

In doing so, we move beyond asking whether “mLearning works” to examine for whom it is effective, under what conditions, and through which mechanisms it supports academic success. These issues are particularly relevant for medical educators who seek to design adaptive digital learning environments that address learner diversity and/or foster engagement within structured curricula. Guided by this perspective, our study aims to provide a deeper understanding of mLearning as an educational practice embedded in dynamic, real-world contexts.

The study, including the use of German national medical licensing exam data, was approved by the local independent review board (Ref. 20230217 01, 2025-80-ka).

Methods

Evaluation Setting and Methodical Approach

We employed an RE design[54] to investigate *what works, for whom, in what context, and how*. Unlike experimental approaches that estimate average effects, RE focusses CMO patterns in authentic educational settings. Accordingly, no randomization or blinding was applied to capture real-world curricular use.

The evaluation was conducted at a German medical school (six-year curriculum). Within the faculty-wide digitalization initiative *WueMedica*, a general mLearning module was developed and implemented across multiple disciplines. In this study, we focused on its use in the fifth-semester course “*Microbiology, Hygiene, and Virology*”. The subject-specific component of the module was delivered via the *eSquirrel* app (eSquirrel GmbH, Vienna, Austria). Alongside the regular curriculum of a voluntary lecture series and mandatory weekly lab sessions, mLearning was offered as an additional learning resource. Teaching staff delivered the standard curriculum independently, and no direct contact occurred between the study team and students.

Intervention

The intervention consisted of a voluntary, app-based course designed to support exam preparation alongside the curriculum. Content was developed by trained student assistants in collaboration with faculty experts and was based on official course materials, the German national competency-based learning objectives catalogue, and standard textbooks. The course was organized into small chapters with interactive tasks (such as multiple-choice, matching, and fill-in-the-blank formats) and supplemented by images and links to relevant textbook chapters. Gamification features, such as point scoring, a leaderboard, push notifications, and an adaptive spaced-repetition algorithm, were included to promote sustained use. Each time a student repeated a learning unit, the system recorded this as progression to the next level (Level 1–4). The levels therefore reflected the number of repetitions completed, not an increasing in task difficulty. The course comprised 88 chapters in 2023 and was expanded to 138 in 2024.

Students accessed the app via the institution’s learning management system. Registration required an official university email address, allowing linkage to student IDs and informed consent. While multiple registrations were technically possible in 2023, they were effectively restricted in 2024 through email verification and ID matching. The app complied with the *General Data Protection*

Regulation standards (European Union) and supported cross-platform, offline access. Data from non-consenting students were excluded from analysis.

Study Design and Data Sources

The evaluation spanned two consecutive summer semesters (2023 and 2024). The study population comprised all students enrolled in the fifth-semester course “*Microbiology, Hygiene, and Virology*”. Eligible participants were defined as formally enrolled in the course and registered for their first attempt at the final oral exam. Repeat examinees or students without formal registration were excluded.

Data were collected from four sources:

1. App usage data (non-reactive), including activity (chapters completed), task success (points earned), and learning efficiency (chapters per unit time).
2. An online questionnaire administered via *EvaSys* (evasys GmbH, Lüneburg, Germany) at the end of the semester, covering demographics (age, sex), usage, and learning behavior, 19 items adapted from published instruments (*UTAUT*[57], *R-SPQ-2F*[58], *Technology Enabled Formative Assessment in Medical Education*[59], *Use of Unsupervised Online Quizzes in Medical Physiology*[60]), and open-ended questions on user satisfaction. The questionnaire was originally developed in German and then translated into English.
3. Final oral exam scores in microbiology in 2024.
4. Oral and written scores from the first national licensing exam (M1, conducted after the fourth semester), serving as contextual data for prior academic performance.

In 2023, 117 students qualified for participation. Based on an expected 80% participation rate, the study was powered to detect a moderate effect size (Cohen $d \approx 0.58$) at $\alpha = .05$ with 80% power.

Data Analysis

Standardized usage and demographic variables (z -scores, calculated as $z = (X - M) / SD$, X as individual value, mean (M), standard deviation (SD)) were entered into a two-step cluster analysis (log-likelihood distance and the Bayesian information criterion) to identify user groups. The aim was to identify user clusters based on app engagement and demographics. Non-users (no completed chapters) were assigned to a separate cluster, and all clusters were compared with respect to survey responses and performance outcomes.

Survey data were analyzed using principal component analysis (PCA) with pairwise deletion for missing values. Components with eigenvalues > 1 were extracted and subjected to oblique (Oblimin) rotation. Reliability was assessed with Cronbach α , and component scores were rescaled to a 5-point Likert scale for easier interpretation. Variance contributions of the final scales were approximated using regression-based scores from the unrotated solution.

All analyses were conducted using *IBM SPSS Statistics* (Version 29.0.0.0; IBM Corp, New York, USA) at a 95% significance level, applying Welch t tests for unequal variances[61,62] and standard t tests when variances were equal, with two-sided P values reported unless otherwise specified. Reporting followed the *CONSORT-EHEALTH* checklist (V.1.6.1) for mobile health interventions[63] and the *RAMESES-II* standards for realist evaluations[64].

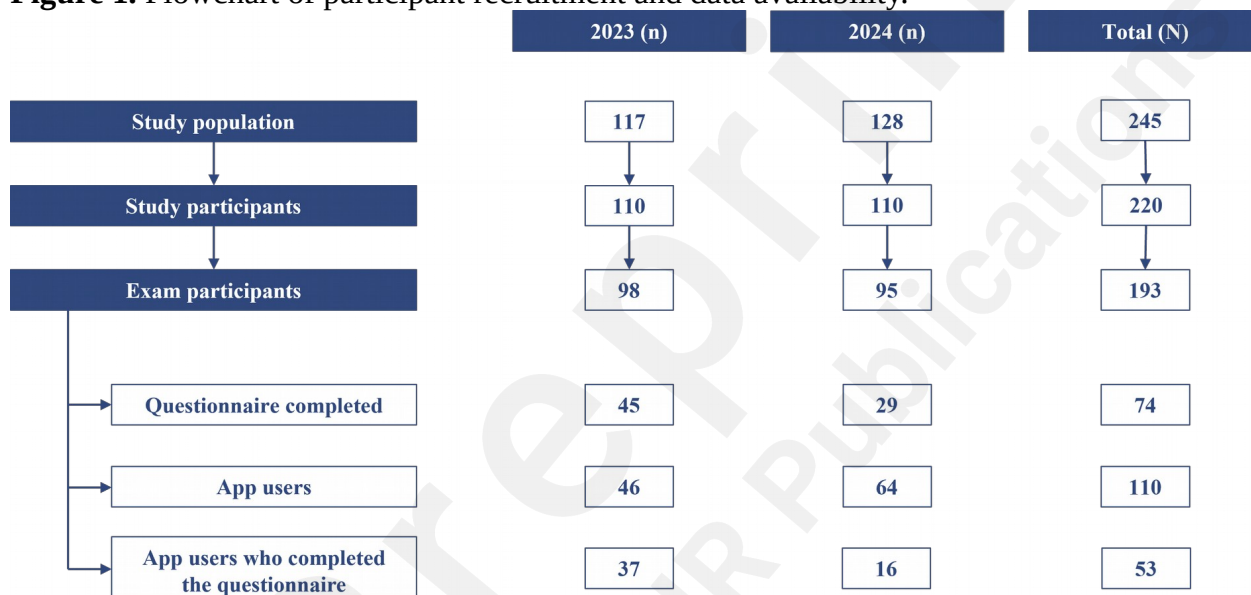
Results

Participant Flow and Sample Characteristics

Participant recruitment and data availability are summarized in Figure 1. An overall recruitment rate of 89.8% (220/245) was achieved, with 110 out of 117 eligible students participating in 2023 (94%) and 110 out of 128 (85.9%) in 2024. Final oral exam results were available for 193 participants (98 in 2023; 95 in 2024), while 27 students (12.3%) were excused from the exam for medical reasons. Half of all participants (110/220) used the mLearning app—46 in 2023 and 64 in 2024—while the remaining 110 formed the non-user comparison group.

The survey produced 74 valid responses (49 in 2023; 25 in 2024), corresponding to response rates of 44.5% and 22.7%. In total, 53 participants (24.1%) both used the app and completed the survey. Demographic characteristics were representative of the eligible population (70% female vs 69%; mean age 23.1 vs 23.9 years). No significant differences were observed between semesters or subgroups. Baseline characteristics are provided in Appendix 1.

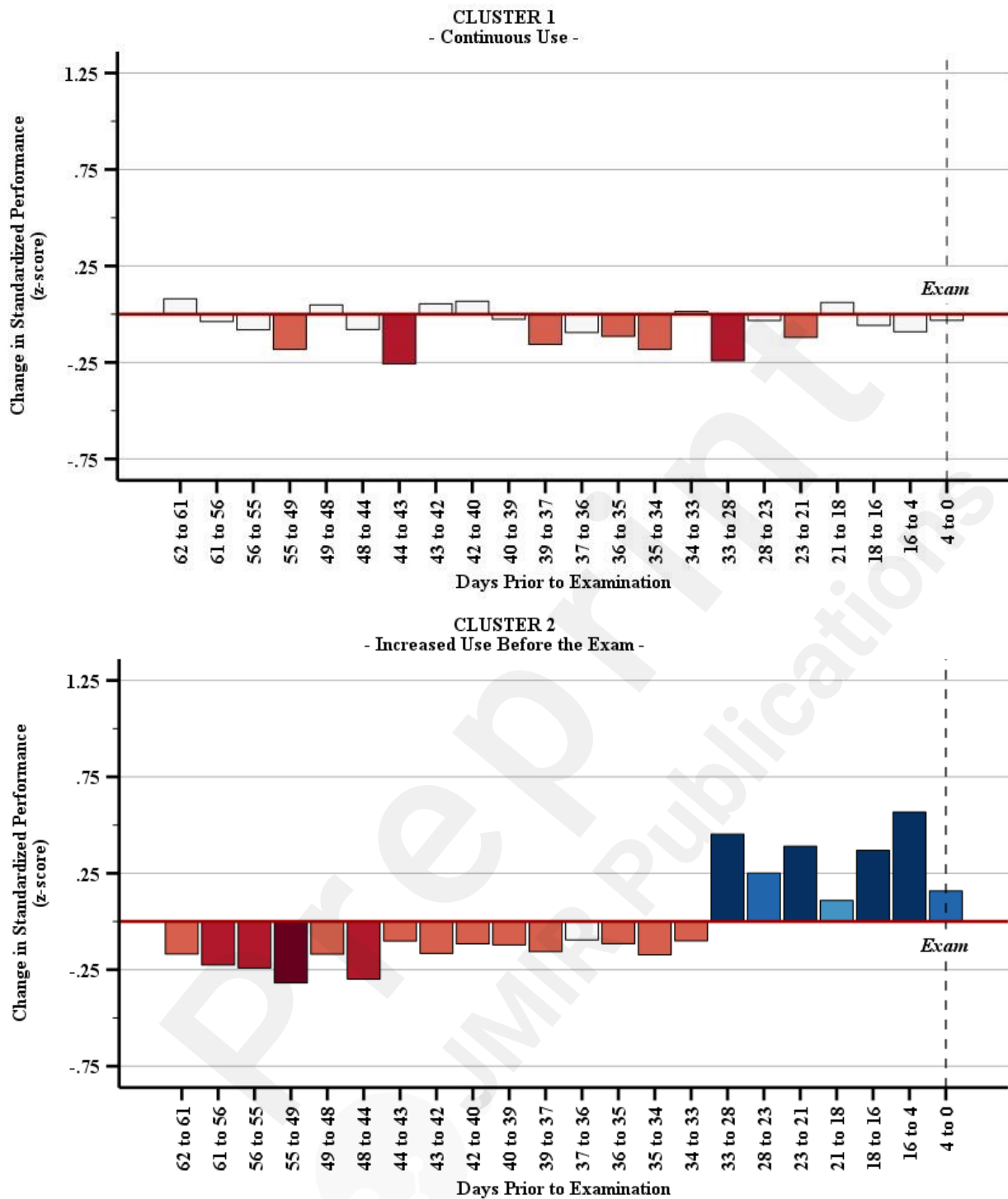
Figure 1. Flowchart of participant recruitment and data availability.

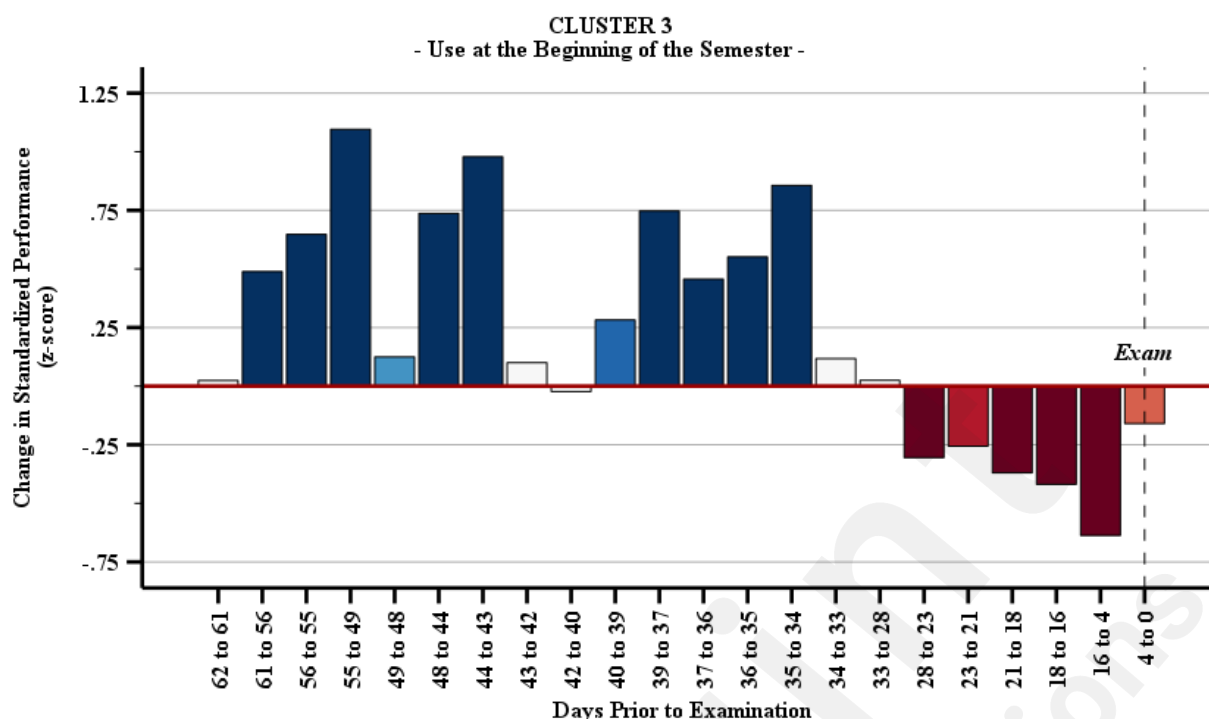


Cluster-Based Usage Profiles

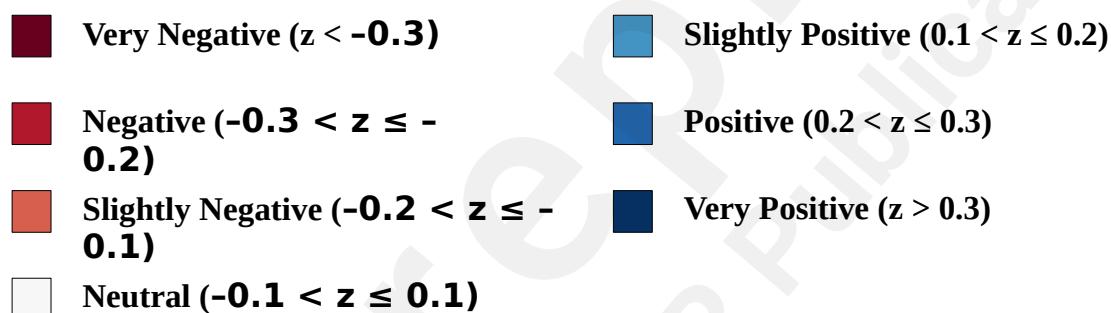
Cluster analysis (average silhouette coefficient $s_c = .40$) of the 110 app users identified three distinct clusters: (1) *Continuous Use*, (2) *Increased Use Before the Exam*, (3) *Use at the Beginning of the Semester*). No significant demographic differences were observed across clusters. Figure 2 presents the z-standardized longitudinal changes in app performance of each cluster throughout the semester, providing the basis for the following detailed description of cluster characteristics:

Figure 2. Longitudinal changes in standardized user-performance across clusters throughout the semester.





Performance relative to the mean (z-score)



The bar charts illustrate changes in standardized performance (z-scores) over the semester. Positive values indicate above-average gains relative to all users; negative values indicate below-average gains. For example, between days 39 and 37 before the exam, Cluster 3 users achieved gains about 0.75 SD above the mean.

- Cluster 1 “Continuous Use” (C1, $n=60$; 54.6%) used the app regularly but completed comparatively few chapters. Users were active on 66.6% of available days yet completed only 26.5% of chapters. Engagement was concentrated at Level 1 (78.4%), and task success was lowest (11.6%). Usage remained consistently low without an increase before the exam.
- Cluster 2 “Increased Use Before the Exam” (C2, $n=31$; 28.2%) combined lower temporal activity (48.1% of days) but high content completion (80.9%). Chapter attempts were more evenly distributed across repetition levels, with moderate success (45.4%). Activity intensified markedly at the end of the semester (exam-oriented), peaking shortly before the exam.
- Cluster 3 “Use at the Beginning of the Semester” (C3, $n=19$; 17.3%) demonstrated the highest overall engagement (87.2% of days) and substantial content completion (67.3% of chapters), with a notable high repetition rate (35.6% at Level 4). Performance was the strongest (54.9%). Engagement peaked out early but declined about four weeks before the exam period.

Oral Exam Performance by Usage Group

The final oral exam performance was analyzed for the 2024 cohort, with grades ranging from 1 (Excellent) to 5 (Failed) (Table 1). The mean oral exam score was $M=2.3$, $SD=1.1$ for app users ($n=64$) compared to $M=2.8$, $SD=1.3$ for non-users ($n=46$). A one-sided Welch t test indicated a statistically significant difference between the two groups ($t_{63.0}=1.90$, $P=.031$). Cluster 1 ($n=34$) achieved $M=2.4$, $SD=1.1$, which did not differ significantly from non-users ($P=.091$). Cluster 2 ($n=20$) obtained the best results with $M=2.1$, $SD=0.9$, significantly outperforming non-users ($P=.008$; one-sided). Cluster 3 ($n=10$) recorded $M=2.5$, $SD=1.3$, with no significant difference compared to non-users ($P=.26$). A higher proportion of non-users (24%) than users (6%) were excused from the exam for medical reasons. No statistically significant differences were observed between the three app-user clusters.

Table 1. Exam scores stratified according to app usage subgroups ($N=110$).

	Total	Non-Users	App Users			
			Total	C1	C2	C3
n	110	46	64	34	20	10
Results M (SD)	2.5 (1.2)	2.8 (1.3)	2.3 (1.1)	2.4 (1.1)	2.1 (0.9)	2.5 (1.3)
Grade 1, n (%)	19 (17)	5 (11)	14 (22)	7 (21)	5 (25)	2 (20)
Grade 2, n (%)	36 (33)	12 (26)	24 (38)	12 (35)	8 (40)	4 (40)
Grade 3, n (%)	21 (19)	7 (15)	14 (22)	8 (24)	4 (20)	2 (20)
Grade 4, n (%)	12 (11)	7 (15)	5 (8)	3 (9)	1 (5)	1 (10)
Failed, n (%)	7 (6)	4 (9)	3 (5)	2 (6)	N/A	1 (10)
Excused, n (%)	15 (14)	11 (24)	4 (6)	2 (6)	2 (10)	N/A

Abbreviations: N/A=not applicable, C1=Cluster 1, C2=Cluster 2, C3=Cluster 3

Questionnaire Results: Principal Component Analysis

Sampling adequacy was acceptable ($KMO=.750$; Bartlett test $P<.001$). PCA initially yielded six component-scales, explaining 74.3% of the total variance. Scales 3-5 were excluded due to low internal consistency ($\alpha=0.46/0.09$) and content overlap. Scale 1 included items relating to both *Learning Outcome* and *User Experience* and was split into two scales in line with the theoretical framework[65,66]. Items 6.1 (*Studying sometimes gives me a sense of deep satisfaction*) and 6.8 (*I will recommend mLearning with WueMedica to others*) were reassigned to the scales *Study Motivation* and *User Experience*, respectively. After manual adjustment, the final three scales—*Learning Outcome* (1a), *User Experience* (1b), and *Study Motivation* (2)—contained no overlapping items and required no further rotation. Their variance contributions were estimated using regression-based scores, jointly explaining 50.2 % of the variance and serving as the basis for subsequent analyses (Table 2):

- The *Learning Outcome* scale (Cronbach $\alpha=0.90$) measured perceived improvements in learning efficiency, comprehension, and long-term retention.
- The *User Experience* scale (Cronbach $\alpha=0.82$) reflected how students evaluated the app's usability, enjoyment, motivational impact, feedback on progress, insight into strengths and weaknesses, curricular relevance, and likelihood of recommending to peers.
- The *Study Motivation* scale (Cronbach $\alpha=0.67$) assessed general study motivation, avoidance of superficial learning strategies, openness to engaging topics, and preference for self-testing.

All scores were coded so that higher values represent stronger agreement with the underlying items. Detailed questionnaire results by usage group are provided in Appendix 2.

Table 2. Principal component analysis results: item loadings and internal consistency.

		Item Loading		
		Learning Outcome	User Experience	Study Motivation
Item				
4.1	mLearning with WueMedica enhances my learning outcomes.	.92		
4.2	mLearning with WueMedica makes my studying more effective.	.92		
4.3	mLearning with WueMedica improves my understanding of the learning topics.	.80		
4.4	mLearning with WueMedica helps me retain what I have learned for a longer period.	.59		
5.1	mLearning with WueMedica is intuitive.		.57	
5.2	mLearning with WueMedica excites me.		.87	
5.3	mLearning with WueMedica motivates me to study more than I did before.		.60	
5.4	mLearning with WueMedica is enjoyable.		.70	
5.5	mLearning with WueMedica provides me with feedback on my learning progress.		.49	
5.6	mLearning with WueMedica helps me identify my strengths and weaknesses in learning.		.46	
5.7	mLearning with WueMedica adds value to the curricular teaching.		.82	
6.1 ^b	Studying sometimes gives me a sense of deep satisfaction.			-.27
6.2	My goal is to pass the course with as little effort as possible.			.53
6.3	I only study the topics that are relevant for the exam.			.73
6.4	Any topic can become interesting to me once I am engaged with it.			-.44
6.5	I learn some things by repeating them until I have memorized them, even if I do not understand them.			.68
6.6	I believe that I can pass most exams by memorizing the key			.68

	points rather than understanding them.			
6.7	I like to test myself on the key topics until I fully understand them.			-.65
6.8	I will recommend mLearning with WueMedica to others.		.90	
Cronbach α		0.90	0.82	0.67
Explained Variance (%)		22.8	13.4	14.0

Questionnaire Results by Usage Group

No statistically significant differences were found between app users and non-users on the scales *Learning Outcome*, *User Experience*, or *Study Motivation*. Item-level analyses, however, revealed consistent advantages for app users, most notably on the enjoyment item.

Cluster 1 did not differ significantly from non-users at the scale level (*Learning Outcome* M=2.6, SD=0.89; *User Experience* M=2.5, SD=0.90; *Study Motivation* M=3.0, SD=0.85. At the item level, only Item 5.4 (enjoyment) was rated significantly more favorably (M=2.1, SD=1.1; $P=.021$; one-sided).

Cluster 2 scored significantly more favorably than non-users on *Learning Outcome* (M=2.0, SD=0.7; one-sided $P=.035$) and *User Experience* (M=2.1, SD=0.70; $P=.034$; one-sided). *Learning Outcome* scores were also significantly higher than those in Cluster 1 ($P=.015$). No significant differences were observed on the *Study Motivation* scale (M=3.1, SD=1.1; $P=.46$). At the item level, Cluster 2 agreed significantly more strongly than non-users on multiple items: Item 4.1 ("mLearning with WueMedica enhances my learning outcomes", $P=.007$), Item 4.4 ("mLearning with WueMedica helps me retain what I have learned for a longer period", $P=.045$), Item 5.4 ($P=.003$), Item 5.6 ("mLearning with WueMedica helps me identify my strengths and weaknesses in learning", $P=.042$), and Item 6.1 ("Studying sometimes gives me a sense of deep satisfaction", $P=.020$). For Item 6.8 ("I will recommend mLearning with WueMedica to others"), scores were also significantly higher than those of non-users ($P=.006$) and Cluster 1 ($P=.014$).

Cluster 3 did not differ significantly from other groups at the scale level (*Learning Outcome* M=2.4, SD=0.74; *User Experience* M=2.3, SD=0.70; and *Study Motivation* M=2.9, SD=1.0. At the item level, only Item 5.4 (enjoyment) was rated significantly more favorably than by non-users (M=2.1, SD=0.9; $P=.012$; one-sided).

Use of Additional Learning Resources

Non-users reported using other digital learning platforms more often than users (60% vs 37%; $P=.044$; one-sided). Across the user clusters, the number of additional resources was highest in Cluster 2 (M=3.6, SD=1.0) and lowest in Cluster 3 (M=3.1, SD=0.9), with the difference demonstrating a statistical trend ($P=.052$; one-sided).

Open-Ended Responses

Open-ended responses were provided across all groups. Among non-users and Cluster 1, the most common themes were time constraints and a perceived misalignment between the app content and the oral exam format. Respondents in this group also noted that multiple-choice tasks were insufficient for preparing for an oral exam.

Participants from Clusters 2 and 3 more frequently highlighted motivational aspects and emphasized the app's contribution to learning microbiology. In these clusters, references to gamification features such as rankings appeared more often (see Appendices 3-5 for details).

Exam Performance

Exam Performance: Users vs Non-Users

Analyses focused on the summer semester of 2024, comprising 35 non-users and 60 app users across the three clusters. The national licensing exam grade (oral and written) served as a baseline indicator of prior academic ability, while the oral microbiology exam grade reflected current performance in the fifth semester. Outcomes according to the groupings are presented in Table 3.

Table 3. Examination performance by usage clusters^a (N=210^b).

			Microbiology	National licensing exam		
		n	Oral exam grade	Overall exam grade	Oral exam grade	Written exam percentile
			Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Eligible population						
		210	2.5 (1.2)	2.6 (0.8)	2.2 (0.8)	74.0 (8.5)
Total participants						
		189	2.5 (1.2)	2.6 (0.8)	2.2 (0.9)	73.9 (8.5)
	Summer 2023					
		94	2.5 (1.2)	2.4 (0.8)	2.1 (0.9)	75.6 (8.8)
	Summer 2024					
		95	2.5 (1.2)	2.7 (0.7)	2.3 (0.8)	72.2 (7.8)
Non-Users						
		35	2.8 (1.3)	2.8 (0.7)	2.3 (0.8)	70.2 (8.1)
Users						
		60	2.3 (1.1)	2.7 (0.7)	2.3 (0.8)	73.4 (7.5)
	Cluster 1					
		32	2.4 (1.1)	2.8 (0.7)	2.4 (0.8)	72.3 (6.8)
	Cluster 2					

		18	2.1 (0.9)	2.5 (0.7)	2.2 (0.8)	73.8 (8.2)
	Cluster 3					
		10	2.5 (1.3)	2.6 (0.4)	2.3 (0.5)	76.6 (8.1)

^aGrades for both microbiology and national licensing exam (M1) range from 1 (Excellent) to 5 (Failed). The written M1 component is reported as percentile rank scores based on the number of points achieved relative to the maximum possible score.

^bData are presented for the 210 participants with available M1 scores; the total study sample comprised 220 individuals.

No significant differences were observed between users and non-users in the overall national licensing exam grade (2.7 vs 2.8; two-sided $P=.37$) or the oral exam grade (2.3 in both groups; two-sided $P=.85$). Users, however, achieved slightly higher written percentile ranks than non-users (73.4 vs 70.2 percentile; $P=.049$).

When comparing performance over time, the results of non-users in the national licensing exam and the microbiology oral exam remained similar ($P=.95$), whereas the results of users improved significantly ($P=.024$). In the oral microbiology exam specifically, non-users' performance declined compared to their national licensing exam (2.3 to 2.8; $P=.029$), while users' performance remained stable.

An interaction between group $\times$ performance change was only close to significance (two-sided $P=.061$). After adjusting for the written national licensing exam grades, ANCOVA revealed significant group $\times$ exam effects for overall grades ($F(1, 92)=3.9$, $P=.050$; $\eta_p^2=0.04$) and for oral grades ($F(1, 92)=3.9$, $P=.030$; $\eta_p^2=0.05$). Written national licensing exam grades significantly predicted microbiology oral exam results ($F(1, 92)=8.0$, $P=.006$ $\eta_p^2=0.08$).

In summary, while users initially outperformed non-users on baseline measures, adjusted analyses suggest that differences in microbiology exam performance were largely explained by prior academic ability, with only limited evidence speaking for any additional benefit of app use.

Exam Performance by Usage Cluster

No significant differences in overall or oral national licensing exam grades were observed across usage clusters. In the written exam, Cluster 3 (76.6 percentile) significantly outperformed non-users (one-sided $P=.022$); while Cluster 2 (73.8 percentile) exhibited a non-significant trend ($P=.068$).

Within clusters, Cluster 1 improved significantly from the overall national licensing exam grade to the final microbiology oral exam (one-sided $P=.048$). Cluster 2 exhibited a similar trend (one-sided $P=.052$), whereas Cluster 3 showed no change. Written-to-oral exam improvements were therefore significant for Clusters 1 and 2, but not for Cluster 3.

A significant interaction was found for oral national licensing exam grades between Cluster 2 and non-users (one-sided $P=.041$). No other significant interactions were observed. After adjusting for written national licensing exam performance (ANCOVA), all cluster-level interactions lost significance ($F(1, 91)=1.5$, $P=.23$; $\eta_p^2=.05$).

Overall, these findings suggest that performance differences between clusters were largely attributable to baseline national licensing exam outcomes. Improvements during the semester were observed mainly in Clusters 1 and 2, but after adjusting for national licensing exam performance, no independent app effect could be confirmed.

Discussion

Integration of Findings within Existing Literature

Established models of mLearning emphasize the critical roles of learning context and individual learner characteristics in shaping engagement and outcomes[17,19,20,24,67]. However, evidence under authentic curricular conditions remains limited. Our study addresses this gap by combining effectiveness and technology acceptance perspectives to identify learner profiles and CMO patterns, thereby clarifying how contextual and individual factors influence mLearning success. Importantly, we did not adopt a randomized controlled trial design. This was a deliberate choice to capture real-world curricular use, not a methodological shortcoming. Allowing participants to regulate their use of mLearning freely, in line with the theory of planned behavior[68] and the concept of self-regulated learning[69,70], was essential to understanding how mLearning operates in practice[71]. Our findings contextualize the “for whom, under what conditions, and how” question posed in the introduction by illustrating that mLearning effectiveness is not universal but varies with learner context and engagement strategy.

Cluster analysis revealed heterogeneous user profiles with distinct engagement patterns and learning outcomes, reinforcing the prior literature emphasizing the personal and individualized nature of mLearning[72]. While mLearning tools can facilitate self-directed learning[73], our results underline that effectiveness depends on learner characteristics and usage strategies. Users in Cluster 2, characterized by strategic, exam-oriented use, reported the highest satisfaction and strongest performance gains. This aligns with previous research based on the *Technology Acceptance Model (TAM)*[74] and the *Unified Theory of Acceptance and Use of Technology (UTAUT)*[57], which highlight affective determinants of technology adoption. That said, our findings demonstrated that enjoyment emerged as a stronger driver of success than perceived usefulness[75–78].

Our findings confirm that mLearning interventions can positively influence academic performance[52]. However, higher-performing students were more likely to adopt the app, suggesting a self-selection bias. This raises concerns that academically weaker students may be less engaged and self-directed learning may not meet their needs[73,79]. This echoes wider discussions of equity in digital health education, in which students with weaker baseline skills or fewer resources risk being left behind[80,81].

Cluster 3 users, engaging early but reducing use before exams, may have shifted to other strategies, while sporadic use in Cluster 1 correlated with minimal benefits. Non-users and Cluster 1 participants frequently doubted the app’s usefulness, consistent with prior studies emphasizing “usefulness” and “attitude” as determinants of adoption[82,83]. Explicit CMO-configurations are provided in detail in the Appendix 6. Although this study is limited to one German medical school, similar usage patterns and contextual challenges have been reported internationally, suggesting that our findings may have broader relevance but require cross-institutional validation.

Methodologically, our study’s strength lies not in re-confirming that mLearning can “work”, a question already well covered, but in integrating effectiveness and acceptance perspectives under authentic curricular conditions. This approach demonstrates that mLearning is not universally “effective”; rather, outcomes depend on the context and user characteristics. To claim otherwise would be equivalent to stating that “medications works” without regard to dosage, adherence, or patient factors. Our findings emphasize that user satisfaction and strategic engagement are critical mechanisms, underscoring the need to tailor mLearning interventions to learner diversity. To support sustainable adoption, institutions should not only provide technical infrastructure but also invest in curricular integration, faculty support, and training for learners. Combining acceptance models (e.g., *TAM*, *UTAUT*) with realist CMO analysis offers a comprehensive heuristic: acceptance explains why learners start using an app, while CMO analysis clarifies how and for whom it succeeds.

Theoretical and Practical Implications

Our findings support a context-sensitive, mechanism-oriented perspective on mLearning effectiveness. Learners with sufficient baseline ability who use app features intensively and strategically, such as spaced repetition[84-86], gamified feedback[87-89], just-in-time learning[90,91], and microlearning[92-94], achieve superior outcomes. This indicates that digital tools should foster not only functional use but also affective engagement to sustain motivation. Given the heterogeneity of learner engagement, mLearning interventions should adapt to diverse learning profiles by fostering autonomy, metacognitive regulation, and enjoyment. The prominent role of enjoyment resonates with research emphasizing that user experience is as important as pedagogy[95,96]. Emerging advances in artificial intelligence, including real-time analytics and adaptive recommendation systems offer opportunities to individualize mLearning[97-100]. Dynamic adjustment of content difficulty and sentiment-based detection of motivational changes could support timely interventions, enhancing both engagement and efficacy. Finally, mLearning enables seamless integration of formal curricular activities with informal, self-directed study, supporting personalized learning trajectories and continuous knowledge acquisition in line with competency-based medical education.

Limitations

Several limitations warrant consideration. Although participant blinding was unfeasible and may introduce expectancy bias, the inclusion of objective app usage tracking enhances ecological validity. The timing of the post-exam survey may have biased responses; future studies should collect feedback prior to the disclosure of grades or results. Controlling for baseline academic achievement strengthened causal inference, but unmeasured factors such as digital literacy, personality, and time management may still confound results. The grading scale was limited; granularity may have restricted sensitivity to subtle performance differences. Subgroup analyses reduced statistical power, highlighting the exploratory nature of these findings. Finally, the reliance on quantitative data constrains interpretive depth. Future studies should adopt mixed-methods designs, integrating log data, surveys, and qualitative approaches (e.g., interviews, think-aloud protocols) to capture learner perspectives and refine CMO configurations.

Future Directions

This study highlights the value of combining global outcome measures (e.g. exam scores) with granular CMO analyses that account for learner characteristics, behavior, and situational factors. Such multi-level approaches clarify not just whether, but why mLearning succeeds or fails in curricular contexts. To improve comparability across studies, consensus on key constructs and standardized definitions (for example, 'engagement,' 'satisfaction,' 'learning gain') is needed. Broader use of frameworks such as *TAM* or *UTAUT* could also support meta-analyses. AI-enabled personalized learning offers opportunities to create adaptive, learner-centered mLearning environments. Future research should investigate how these innovations affect diverse learner groups, including those with lower baseline performance. Overall, our findings suggest that when baseline ability and strategic app use align with just-in-time exam preparation (*Context*), motivational, self-regulatory, and feedback processes are activated (*Mechanism*), leading to improved exam outcomes (*Outcome*).

Funding

This study was conducted within the *WueMedica* project and supported by *WueDive – Innovation in Education, an intramural funding initiative of the University of Würzburg (Germany)*.

Acknowledgments

We acknowledge Marc Appel for his administrative coordination and support in implementing the *WueMedica* project. We thank Carsten Scheller for his support in reviewing the app content in *WueMedica*. We value the organizational support by Waltraud Flederer to support the successful execution of the project. Finally, we thank Andrew Entwistle for the linguistic and grammatical revision of the manuscript as well as for his valuable suggestions for semantic improvements. No generative AI was used in writing the final manuscript.

Author's Contributions

Conceptualization: SK, JB, JPHH

Data curation: JB, JPHH

Formal analysis: JB, JPHH

Funding acquisition: SK

Investigation: JPHH

Methodology: SK, JB, JPHH

Project administration: SK

Resources: SK, TTL

Software: JB

Supervision: SK

Validation: SK, TTL, JB, JPHH

Visualization: JPHH

Writing – original draft: JPHH

Writing – review & editing: SK, JB, JPHH

Conflicts of Interest

None declared.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Abbreviations

AI: artificial intelligence

C1: Cluster 1

C2: Cluster 2

C3: Cluster 3

CMO: context-mechanism-outcome configuration

EFA: exploratory factor analysis
eLearning: electronic learning
GDPR: General Data Protection Regulation (European Union)
ICF: Informed Consent Form
KMO: Kaiser-Meyer-Olkin Measure of Sampling Adequacy
LMS: Learning management system
LO: Learning Outcome
M: Mean
M1: First Medical Licensing Examination (Germany)
mLearning: mobile learning
N/A: not applicable
n: sub-cohort count
N: total cohort count
P: P value
PCA: principal component analysis
RE: realist evaluation
RCT: randomized controlled trial
UE: User Experience
SM: General Study Motivation
R-SPQ-2F: The Revised Two Factor Study Process Questionnaire
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology
y: years

References

1. Grant MM. Difficulties in defining mobile learning: analysis, design characteristics, and implications. *Education Tech Research Dev* 2019;67(2):361-388. doi:10.1007/s11423-018-09641-4
2. Crompton H. A Historical Overview of M-Learning: Toward Learner-Centered Education. In: Berge ZL, Muilenburg LY, editors. *Handbook of Mobile Learning*. New York, NY, USA: Routledge; 2013. ISBN:9780415503693. p. 45–218.
3. Seipold J. Aus der Geschichte des mobilen Lernens: Strömungen, Trends und White Spaces. In: Witt C de, Gloerfeld C, editors. *Handbuch Mobile Learning*. 1. Auflage. Wiesbaden, Heidelberg: Springer VS; 2018. ISBN:3658191228. p. 13–42.
4. Parzl R, Bannert M. Mobile Learning – Begriff, Modelle, Forschung: Pädagogisch-psychologische und didaktische Aspekte des M-Learning im formellen Bildungskontext. *eleed* 2013(9).doi:10.1016/j.trf.2012.10.001
5. Frohberg D. *Mobile Learning: Dissertation der wirtschaftswissenschaftlichen Fakultät der Universität Zürich* 2008.
6. Winters N. What is mobile learning? In: *Big Issues in Mobile Learning: Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative*. Nottingham, United Kingdom; 2006. p. 5–9.
7. Witt C de, Sieber A, editors. *Mobile Learning: Potenziale, Einsatzszenarien und Perspektiven des Lernens mit mobilen Endgeräten*. Dordrecht: Springer VS; 2013. ISBN:9783531194837.
8. West M, Vosloo S. UNESCO policy guidelines for mobile learning. Paris: United Nations Educational, Scientific and Cultural Organization; 2013. ISBN:978-92-3-001143-7.
9. Ellaway RH, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. *Med Teach* 2008;30(5):455-473. PMID:18576185

10. Ally M, editor. Mobile learning: Transforming the delivery of education and training. Edmonton Alta.: AU Press; 2009. ISBN:9781897425442.
11. Ally M, McGreal R, Lin F, Woo B. An intelligent agent for adapting and delivering electronic course materials to mobile learners. MLearn 2005.
12. Traxler J. Defining Mobile Learning. In: Isaías P, Borg C, Kommers P, Bonanno P, editors. Proceedings of the IADIS International Conference Mobile Learning 2005: IADIS; 2005. ISBN:972-8939-02-7. p. 261–266.
13. Danish J, Hmelo-Silver CE. On activities and affordances for mobile learning. *Contemp Educ Psychol* 2020;60:101829. doi:10.1016/j.cedpsych.2019.101829
14. Kukulska-Hulme A. Mobile learning as a catalyst for change. *Online J Distance Educ Elearn* 2010;25(3):181-185. doi:10.1080/02680513.2010.511945
15. Traxler J. Defining, Discussing, and Evaluating Mobile Learning: The moving finger writes and having writ... *International Review of Research in Open and Distance Learning* 2007;8(2). doi:10.19173/irrodl.v8i2.346
16. Sharples M, Taylor J, Vavoula G. A Theory of Learning for the Mobile Age. In: Haythornthwaite CA, Andrews R, Fransman J, Meyers EM, editors. *The Sage handbook of E-learning research*. Second edition. Los Angeles: SAGE Reference; 2016. ISBN:9781473902329. p. 63–81.
17. Koole ML. A Model for Framing Mobile Learning. In: Ally M, editor. Mobile learning: Transforming the delivery of education and training. Edmonton Alta.: AU Press; 2009. ISBN:9781897425442. p. 25–47.
18. Park Y. A Pedagogical Framework for Mobile Learning: Categorizing Educational Applications of Mobile Technologies into Four Types. *International Review of Research in Open and Distance Learning* 2011;22(2):78-102. doi:10.19173/irrodl.v12i2.791
19. Kearney M, Schuck S, Burden K, Aubusson P. Viewing mobile learning from a pedagogical perspective. *Research in Learning Technology* 2012;20(1). doi:10.3402/rlt.v20i0.14406
20. Petit T, Lacerda Santos G. Mobile Learning: An Ergonomic Alternative for Long-Awaited Educational Changes? In: Bayyurt Y, Kalz M, Specht M, editors. *Mobile as Mainstream - Towards Future Challenges in Mobile Learning: 13th World Conference on Mobile and Contextual Learning, mLearn 2014, Istanbul, Turkey, November 3-5, 2014. Proceedings*. 1st ed. 2014. Cham: Springer International Publishing; Imprint: Springer; 2014. ISBN:9783319134161. p. 1–14.
21. Parsons D, Palalas A, Nikou S, Rodulfo S. Mobile learning frameworks and pedagogy: A systematic review. *Euro J of Education* 2024;59(2). doi:10.1111/ejed.12601
22. McQuiggan S, Kosturko L, McQuiggan J, Sabourin J. *Mobile Learning: A Handbook for Developers, Educators, and Learners*. Cary, North Carolina, USA: SAS Institute Inc; 2015. ISBN:9781118894309
23. Kukulska-Hulme A, Traxler J, editors. *Mobile Learning: A handbook for educators and trainers*. Abingdon: Routledge; 2005. ISBN:9780415357401
24. Sharples M, Taylor J, Vavoula G. Towards a Theory of Mobile Learning. In: van der Merwe H, Brown T, editors. *Mobile technology: The future of learning in your hands: mLearn 2005*; 2005. ISBN:1-86854-630-6. p. 23–27.
25. Dalloz X. With the Internet of Augmented Me, a New Wave of Innovation is Coming, Which Will Change Everything. In: Duvaut P, Dalloz X, Menga D, Koehl F, Chriqui V, Brill J, editors. *Internet of Augmented Me. I.AM: Empowering Innovation for a New Sustainable Future*. London: Wiley-ISTE; 2020. ISBN:978-1-78630-508-4. p. 1–36.
26. Kakiyama M, Sorensen C. Mobility: an extended perspective. In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences: IEEE Comput. Soc*; 2002. ISBN:0-7695-1435-9. p. 1756–1766.
27. Traxler J. Current State of Mobile Learning. In: Ally M, editor. Mobile learning:

- Transforming the delivery of education and training. Edmonton Alta.: AU Press; 2009. ISBN:9781897425442. p. 9–24.
28. Stoller-Schai D. Mobiles Lernen: die Lernform des Homo Mobilis. In: Hohenstein A, Wilbers K, editors. *Handbuch E-Learning: Expertenwissen aus Wissenschaft und Praxis - Strategien, Instrumente, Fallstudien*. Wolters Kluwer; 2010. ISBN:9783871562983
 29. Badillo P-Y, Tarrier F. Mobilité et ubiquité dans le futur: vers le nomadisme numérique. *Cahier N°1 de l'ANR* 2009.
 30. Sharples M. Big Issues in Mobile Learning: Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative. Nottingham, United Kingdom: University of Nottingham; 2006.
 31. Dewey J. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: The Macmillan Company; 1916.
 32. Bartelsen J. *mobile learning: Ein Überblick*. Elmshorn: Hochschule der Wirtschaft; 2011. Arbeitspapiere der Nordakademie; 6.
 33. Laouris Y, Eteokleous N. We need an educationally relevant definition of mobile learning. In: *Proceedings of the 4th World Conference on Mobile Learning 2005*. p. 290–294.
 34. O'Malley C, Vavoula G, Glew J, Taylor J, Sharples M, Lefrere P, Lonsdale P, Naismith L, Waycott J. *MOBILearn Pedagogical Methodologies and Paradigms: Guidelines for learning/teaching/tutoring in a mobile environment* 2005.
 35. Chat M, Zghida N, Ouaili AA, Youlyouz A, Amiri EM, Janati-Idrissi R. Towards a new definition of e-learning and m-learning. *Multidisciplinary Reviews* 2025;8(9):2025281. doi:10.31893/multirev.2025281
 36. Bullock A, Webb K. Technology in postgraduate medical education: a dynamic influence on learning? *Postgrad Med J* 2015;91(1081):646-650. PMID:26341127
 37. Garzón J, Kinshuk, Burgos D, Tlili A. Advantages and challenges associated with mobile learning in education: a systematic literature review. *J. Comput. Educ.* 2024. doi:10.1007/s40692-024-00342-x
 38. Crompton H, Burke D. The use of mobile learning in higher education: A systematic review. *Comput Educ* 2018;123:53-64. doi:10.1016/j.compedu.2018.04.007
 39. Kho A, Henderson LE, Dressler DD, Kripalani S. Use of handheld computers in medical education. A systematic review. *J Gen Intern Med* 2006;21(5):531-537. PMID:16704405
 40. Fischer S, Stewart TE, Mehta S, Wax R, Lapinsky SE. Handheld computing in medicine. *J Am Med Inform Assoc* 2003;10(2):139-149. PMID:12595403
 41. Chandran VP, Balakrishnan A, Rashid M, Pai Kulyadi G, Khan S, Devi ES, Nair S, Thunga G. Mobile applications in medical education: A systematic review and meta-analysis. *PLoS One* 2022;17(3):e0265927. PMID:35324994
 42. Dimond R, Bullock A, Lovatt J, Stacey M. Mobile learning devices in the workplace: 'as much a part of the junior doctors' kit as a stethoscope'? *BMC Med Educ* 2016;16(1):207. PMID:27530343
 43. Traxler J. What killed the mobile learning dream? 2016 URL: <https://medium.com/@Jisc/what-killed-the-mobile-learning-dream-8c97cf66dd3d> [accessed 2025-10-06].
 44. Karimi S. Do learners' characteristics matter? An exploration of mobile-learning adoption in self-directed learning. *Comput Human Behav* 2016;63:769-776. doi:10.1016/j.chb.2016.06.014
 45. Lewis KO, Popov V, Fatima SS. From static web to metaverse: reinventing medical education in the post-pandemic era. *Ann Med* 2024;56(1):2305694. PMID:38261592
 46. Car J, Ong QC, Erlikh Fox T, Leightley D, Kemp SJ, Švab I, Tsoi KKF, Sam AH, Kent FM, Hertelendy AJ, Longhurst CA, Powell J, Hamdy H, Nguyen HVQ, Aoun Bahous S, Wang M, Baumgartner M, Mahendradhata Y, Popovic N, Khong AWH, Prober CG, Atun R, Bekele

- Zerihun A, Poncette A-S, Molina AJR, Ferreira AVL, Fajkic A, Kaushal A, Farmer AJ, Lane AS, Kononowicz AA, Bhongir AV, Alayande BT, Bene BA, Dameff CJ, Hallensleben C, Back DA, Hawezzy DJ, Tulantched DSM, Kldiashvili E, Achampong EK, Ramachandran G, Hauser G, Grove J, Cheung JPY, Imaralu JO, Sotunsa JO, Bulnes Vides JP, Lawrence KS, Agha-Mir-Salim L, Saba L, Zhang L, Elfiky MMA, Hesseling MW, Guppy MP, Phatak MS, Al Saadoon MAA, Lai NM, Chavannes NH, Kimberger O, Povia P, Goh P-S, Grainger R, Nannan Panday RS, Forsyth R, Vento S, Lee SY, Yadav SK, Syed-Abdul S, Appenzeller S, Denaxas S, Garba SE, Flügge T, Bokun T, Dissanayake VHW, Ho V, Obadiel YA. The Digital Health Competencies in Medical Education Framework: An International Consensus Statement Based on a Delphi Study. *JAMA Netw Open* 2025;8(1):e2453131. PMID:39888625
47. Wissenschaftsrat. Neustrukturierung des Medizinstudiums und Änderung der Approbationsordnung für Ärzte: Empfehlungen der Expertenkommission zum Masterplan Medizinstudium 2020. Köln: Wissenschaftsrat; 2018.
 48. Gloerfeld C. Mobile Learning - was ist eigentlich der Kern: Eine Quintessenz aus Theorie, Modellen und Konzepten. In: Witt C de, Gloerfeld C, editors. *Handbuch Mobile Learning*. 1. Auflage. Wiesbaden, Heidelberg: Springer VS; 2018. ISBN:3658191228. p. 257–282.
 49. High Level Group on the Modernisation of Higher Education. Report to the European Commission on Improving the Quality of Teaching and Learning in Europe's Higher Education Institutions. Luxembourg; 2013.
 50. Khurana MP, Raaschou-Pedersen DE, Kurtzhals J, Bardram JE, Ostrowski SR, Bundgaard JS. Digital health competencies in medical school education: a scoping review and Delphi method study. *BMC Med Educ* 2022;22(1):129. PMID:35216611
 51. Barradell S, Bell A, Thomson K, Hughes J. Students' informal learning interactions in health professions education: insights from a qualitative synthesis 1. *Higher Education Research & Development* 2025;44(2):322-337. doi:10.1080/07294360.2024.2393125
 52. Bashir F, Warraich NF, Malik A. Paradigm Shift of Health Professions Education Through Mobile Devices: A Literature Review of Reviews. *Med Sci Educ* 2025;35(2):1137-1147. PMID:40352993
 53. Parsons D. A Mobile Learning Overview by Timeline and Mind Map. *International Journal of Mobile and Blended Learning* 2014;6(4):1-21. doi:10.4018/ijmbl.2014100101
 54. Pawson R, Tilley N. *Realistic evaluation*. London, Thousand Oaks Calif.: Sage; 1997. ISBN:0761950087.
 55. Lemire S, Kwako A, Nielsen SB, Christie CA, Donaldson SI, Leeuw FL. What Is This Thing Called a Mechanism? Findings From a Review of Realist Evaluations. *New Directions for Evaluation* 2020;2020(167):73-86. doi:10.1002/ev.20428
 56. Nielsen SB, Lemire S. Nothing as practical as an analytical strategy in realist evaluation. In: Manzano A, Williams E, editors. *Realist Evaluation*. London: Routledge; 2024. ISBN:9781003457077. p. 40–56.
 57. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 2003;27(3):425-478. doi:10.2307/30036540
 58. Biggs J, Kember D, Leung DY. The revised two-factor Study Process Questionnaire: R-SPQ-2F. *Br J Educ Psychol* 2001;71(1):133-149. PMID:11307705
 59. Youhasan P, Raheem S. Technology Enabled Formative Assessment in Medical Education: A Pilot Study through Kahoot. *EIMJ* 2019;11(3):23-29. doi:10.21315/eimj2019.11.3.3
 60. Kibble J. Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Adv Physiol Educ* 2007;31(3):253-260. PMID:17848591
 61. Rasch D, Kubinger KD, Moder K. The two-sample t test: pre-testing its assumptions does not pay off. *Stat Papers* 2011;52(1):219-231. doi:10.1007/s00362-009-0224-x

62. Rasch D, Teuscher F, Guiard V. How robust are tests for two independent samples? *J Stat Plan Inference* 2007;137(8):2706-2720. doi:10.1016/j.jspi.2006.04.011
63. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res* 2011;13(4):e126. PMID:22209829
64. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. *BMC Med* 2016;14(1):96. PMID:27342217
65. O'Rourke N, Hatcher L. *A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling*. Second Edition. Cary, North Carolina: SAS Institute Inc; 2013.
66. Savalei V, Huang M. Fit indices are insensitive to multiple minor violations of perfect simple structure in confirmatory factor analysis. *Psychological Methods* 2025. doi:10.1037/met0000718
67. Terras MM, Ramsay J. The five central psychological challenges facing effective mobile learning. *Brit J Educational Tech* 2012;43(5):820-832. doi:10.1111/j.1467-8535.2012.01362.x
68. Ajzen I. *From Intentions to Actions: A Theory of Planned Behavior*. In: Kuhl J, editor. *Action Control: From Cognition to Behavior*. Berlin, Heidelberg: Springer; 1985. ISBN:9783642697463. p. 11–39.
69. Zimmerman BJ. Becoming a Self-Regulated Learner: An Overview. *Theory Pract* 2002;41(2):64-70. doi:10.1207/s15430421tip4102_2
70. Chi NY, Huyen NT. Self-Regulation of Learning in Higher Education Mobile Learning Landscape: A Review. *VNU Journal of Science: Education Research* 2023. doi:10.25073/2588-1159/vnuer.5207
71. Cheon J, Lee S, Crooks SM, Song J. An investigation of mobile learning readiness in higher education based on the theory of planned behavior. *Comput Educ* 2012;59(3):1054-1064. doi:10.1016/j.compedu.2012.04.015
72. Sharples M. The design of personal mobile technologies for lifelong learning. *Comput Educ* 2000;34:177-193.
73. Bhat PP, Rajashekar B, Kamath U. Perspectives on Self-Directed Learning — the Importance of Attitudes and Skills. *Bioscience Education* 2007;10(1):1-3. doi:10.3108/beej.10.c3
74. Davis FD. *A Technology Acceptance Model For Empirically Testing New End-User Information Systems: Theory and Results*. Massachusetts Institute of Technology; 1985.
75. Kucuk S, Baydas Onlu O, Kapakin S. A Model for Medical Students' Behavioral Intention to Use Mobile Learning. *J Med Educ Curric Dev* 2020;7:2382120520973222. PMID:33313399
76. Chao C-M. Factors Determining the Behavioral Intention to Use Mobile Learning: An Application and Extension of the UTAUT Model. *Front Psychol* 2019;10:1652. PMID:31379679
77. Granić A, Marangunić N. Technology acceptance model in educational context: A systematic literature review. *Brit J Educational Tech* 2019;50(5):2572-2593. doi:10.1111/bjet.12864
78. Nabipour Sanjebad N, Shrestha A, Shahid P. The Impact of Personality Traits Towards the Intention to Adopt Mobile Learning. In: Sharma SK, Dwivedi YK, Metri B, Rana NP, editors. *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation*. Vol. 618. Cham: Springer International Publishing; 2020. ISBN:978-3-030-64860-2. p. 182–193.
79. Golenhofen N, Heindl F, Grab-Kroll C, Messerer DAC, Böckers TM, Böckers A. The Use of a Mobile Learning Tool by Medical Students in Undergraduate Anatomy and its Effects on Assessment Outcomes. *Anat Sci Educ* 2020;13(1):8-18. PMID:30913369
80. Han SP, Kumwenda B. Bridging the digital divide: Promoting equal access to online learning for health professions in an unequal world. *Med Educ* 2025;59(1):56-64. PMID:38860836
81. König LM, Krukowski RA, Kuntsche E, Busse H, Gumbert L, Gemesi K, Neter E, Mohamed

- NF, Ross KM, John-Akinola YO, Cooper R, Allmeta A, Silva AM, Forbes CC, Western MJ. Reducing intervention- and research-induced inequalities to tackle the digital divide in health promotion. *Int J Equity Health* 2023;22(1):249. PMID:38049789
82. Thi TDP, Pham VK, Duong, Nam, Tien. Understanding M-learning app adoption: an integrated model for college students in Vietnam. *Interactive Learning Environments* 2025;1-23. doi:10.1080/10494820.2025.2488144
83. Wu H, Que X, Pan L. Analysing students' aspiration factors that impact actual use of mobile learning systems: a two-stage SEM-ANN approach. *Humanit Soc Sci Commun* 2025;12(1). doi:10.1057/s41599-025-04717-y
84. Ebbinghaus H. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. 1st ed. Leipzig: Duncker & Humblot; 1885.
85. Cepeda N, Pashlar H, Vul E, Wixted J, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 2006;132(3):354-380. doi:10.1037/0033-2909.132.3.354
86. Schimanke F. The Impact of Spaced Repetition Learning on the Learning Success in Mobile Learning Games. In: 2021 IEEE International Symposium on Multimedia (ISM): IEEE; 2021. ISBN:978-1-6654-3734-9. p. 275–280.
87. Ratinho E, Martins C. The role of gamified learning strategies in student's motivation in high school and higher education: A systematic review. *Heliyon* 2023;9(8):e19033. PMID:37636393
88. Krishnamurthy K, Selvaraj N, Gupta P, Cyriac B, Dhurairaj P, Abdullah A, Krishnapillai A, Lugova H, Haque M, Xie S, Ang E-T. Benefits of gamification in medical education. *Clin Anat* 2022;35(6):795-807. PMID:35637557
89. Su C-H, Cheng C-H. A mobile gamification learning system for improving the learning motivation and achievements. *Computer Assisted Learning* 2015;31(3):268-286. doi:10.1111/jcal.12088
90. Breitwieser J, Neubauer AB, Schmiedek F, Brod G. Realizing the potential of mobile interventions for education. *NPJ Sci Learn* 2024;9(1):76. PMID:39668147
91. Riel M. Education in the 21st Century: Just-in-Time Learning or Learning Communities. Emirates Center for Strategic Studies and Research 1998:137-160.
92. HajiAbad FA, Keshmiri F, Jabinian F, Shiryazdi SM. Micro-learning and Mindtools in Mobile-Learning (M-Learning): effects on knowledge and self-efficacy in laparoscopy education for surgical nursing students. *BMC Nurs* 2025;24(1):202. PMID:39984878
93. Thillainadesan J, Le Couteur DG, Haq I, Wilkinson TJ. When I say ... microlearning. *Med Educ* 2022;56(8):791-792. PMID:35654438
94. Hug T. Mikrolernen: konzeptionelle Überlegungen und Anwendungsbeispiele. In: Herzig B, Meister DM, Moser H, Niesyto H, editors. *Jahrbuch Medienpädagogik 8: Medienkompetenz und Web 2.0*. 1. Aufl. Wiesbaden: VS Verl. für Sozialwiss; 2010. ISBN:9783531921358. p. 221–238.
95. Kim C-M, Pekrun R. Emotions and Motivation in Learning and Performance. In: Spector JM, editor. *Handbook of research on educational communications and technology*. Fourth edition. New York: Springer; 2014. ISBN:9781461431848. p. 65–76.
96. Smith CA, Lazarus RS. Emotion and Adaptation. *Contemporary Sociology* 1992;21(4):522. doi:10.2307/2075902
97. Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Med Educ* 2025;25(1):129. PMID:39871336
98. Zhu Y, Dai W, Kang Q. The analysis of artificial intelligence-based mobile learning in students' open teaching recommendation system based on deep learning. *Sci Rep* 2025;15(1):21927. PMID:40595317

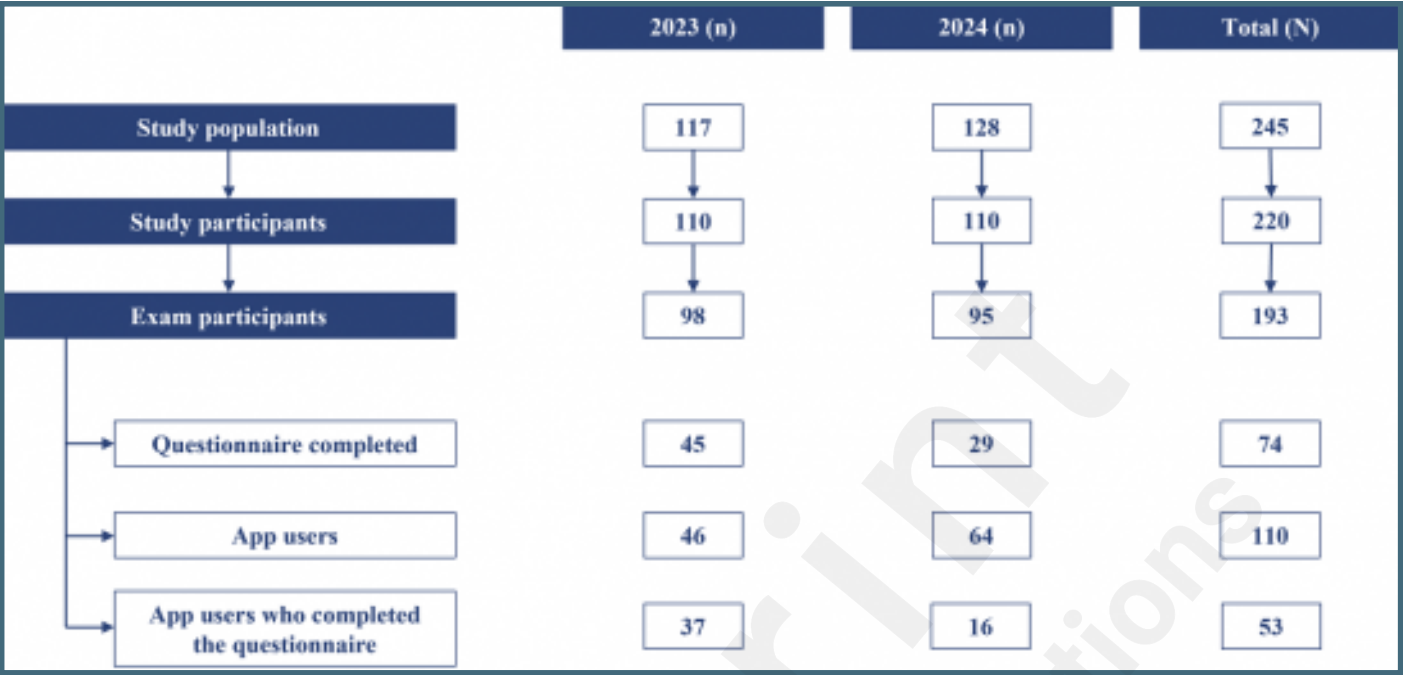
99. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. *Acad Med* 2024;99(1):22-27. PMID:37651677
100. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial Intelligence in Undergraduate Medical Education: A Scoping Review. *Acad Med* 2021;96(11S):S62-S70. PMID:34348374



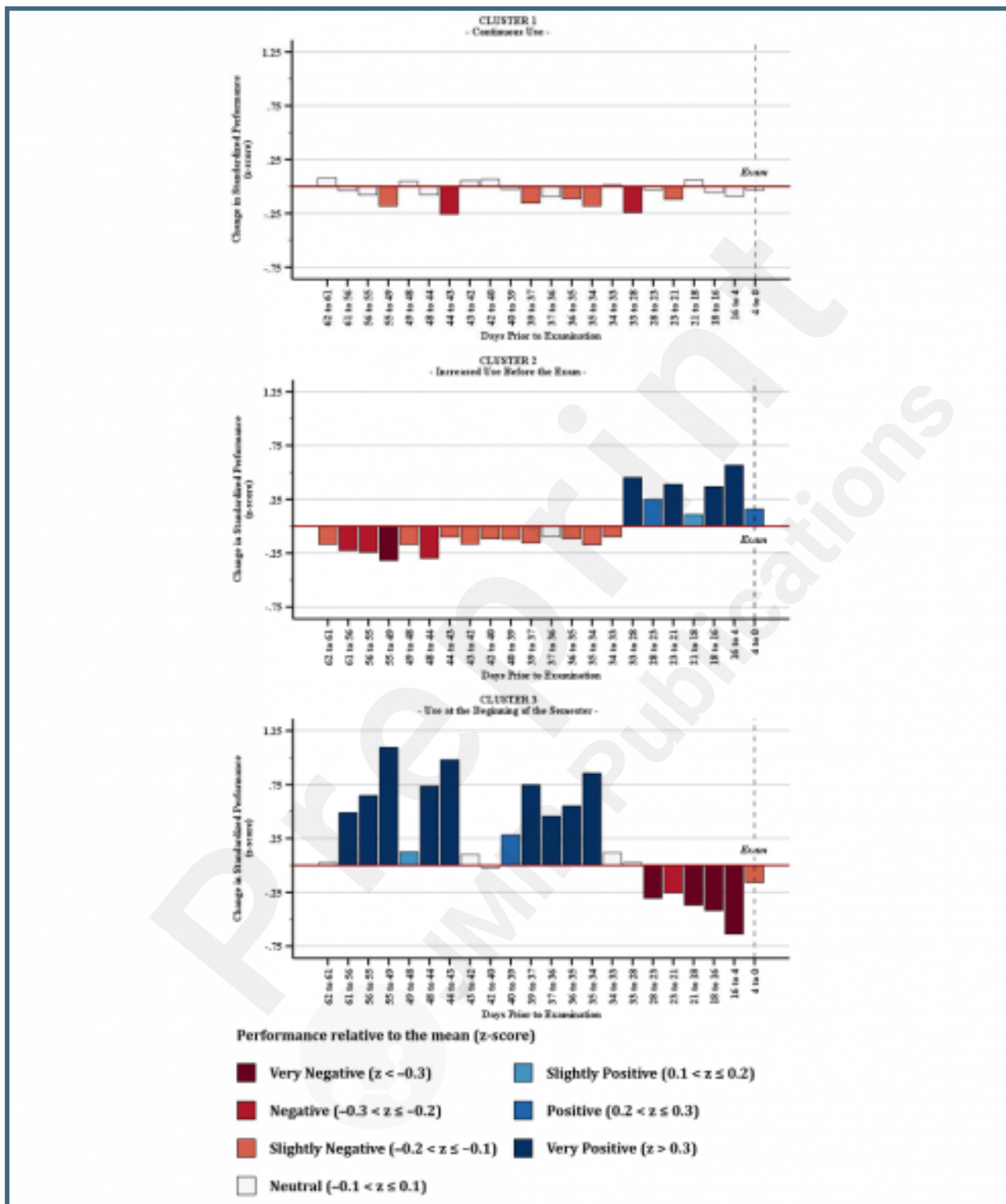
Supplementary Files

Figures

Flowchart of participant recruitment and data availability.



Longitudinal changes in standardized user-performance across clusters throughout the semester.



Multimedia Appendixes

Baseline characteristics of participants stratified by study participation stage (N=220)a.

URL: <http://asset.jmir.pub/assets/9e6585b026d9f1b01c0083f7a6f75287.docx>

Questionnaire results: component scores and item-level data by user status and usage clusters (N=74a).

URL: <http://asset.jmir.pub/assets/864c5d0779f08c999b12e7eb3eedc44b.docx>

Q_7.1: What did you particularly like about using the WueMedica course.

URL: <http://asset.jmir.pub/assets/194e1f796d483d9068d77adf09823d25.docx>

Q_7.2 Where do you see room for improvement in the WueMedica course.

URL: <http://asset.jmir.pub/assets/c40d5613c1a130225aa717af182c2f60.docx>

Q_7.3: What else would you like to tell us about mobile learning...

URL: <http://asset.jmir.pub/assets/d104145f51835442e706736c2b608007.docx>

CMO-configurations.

URL: <http://asset.jmir.pub/assets/a5fc5275025645c866c9c5b27b10bfb6.docx>