A Comprehensive Usability Measurement Tool for m-Learning Applications

Christian X. Navarro-Cota, Ana I. Molina, Miguel A. Redondo, and Carmen Lacave

***Abstract*—*Contribution:* This article describes the process used to create a questionnaire to evaluate the usability of mobile learning applications (CECAM). The questionnaire includes specific questions to assess user interface usability and pedagogical usability.**

***Background:* Nowadays, mobile applications are expanding rapidly and are commonly used in educational institutions to support the learning and teaching process. But the possible deficient usability could decrease the utility of learning activities and the student's motivation. Therefore, careful planning and design by the developer are required, along with a usability evaluation of the applications.**

***Research Questions:* How could an instrument be developed to evaluate the usability of m-learning applications that combine technical and pedagogical aspects? How can the quality of the developed instrument be determined?**

***Methodology:* A structured questionnaire was created like a measuring tool to evaluate and design m-learning applications. Different statistical techniques, including reliability and validity assessments, were employed to evaluate the quality of the instrument, which is determined through the calibration of the CECAM survey.**

***Findings:* After the validity analysis of the questionnaire, a scale with 56 items was obtained, with an alpha reliability coefficient of 0.911 (an excellent measuring scale). It pretends to be used by teachers to design or evaluate m-learning applications, improve their usability, and enhance the students' learning experience.**

***Index Terms*—Evaluation, guidelines, heuristics, instrument development, m-learning, reliability, survey, usability, validity.**

I. INTRODUCTION

THE USE of *mobile devices* (smartphones and especially tablets) to support teaching and learning activities has gained importance and has been consolidated over the last few years [1]. This has led to the popularization of different terms that have emerged to refer to this new scenario [2]: m-learning, ubiquitous learning, seamless learning, blended learning, or smart education.

However, the most widely used and widespread term is *mobile learning* (or m-learning) [3]. Although the concept of *m-learning* includes a large number of variants, with no consensus on its definition [4], one of the most widely accepted definitions is that of [5]: "*the processes of coming to know through conversations across multiple contexts amongst people and personal interactive technologies*." In [3] it is defined as "*learning across multiple contexts, through social and content interactions, using personal electronic devices*." This last definition is more focused on the students and their learning process.

Since the concept of m-learning has become widespread, several authors have directed their studies to demonstrate its benefits [6], identify the new challenges that arise [7], as well as discuss its advantages and disadvantages [8]. In their works, these authors highlight several factors that can be considered determinants for the advancement of m-learning:

1) An adequate technological and pedagogical integration is required.
2) Facilitates access to information at any time and in any place, breaking down geographical and time barriers.
3) Allows the adaptation of interfaces, contents, methodologies, and activities to the individual differences of students, in addition to generating personalized analysis and feedback.
4) Promote communication and knowledge sharing, fostering collaborative learning.
5) Contribute to enhancing student motivation, especially when it is used to implement innovative methodologies, such as those that exploit the use of gamification.

A fundamental feature to make the above factors contribute positively is the *usability* of the tools, since this aspect has a critical impact on the performance of learning activities [9], [10], [11]. In the field of Human-Computer Interaction, in-depth work has been done on the usability of software applications (technical usability); and techniques and tools have been developed to contemplate its requirements in the design phases and posterior evaluation [12], [13]. However, the context and support for m-learning [14]define a more complex and dynamic scenario due to the specific characteristics of the devices used (small screens, limited input capabilities, mobility, etc.) [15], [16].

The proposal of guidelines for the specific development of e-learning applications has been a focus of research by

Manuscript received 9 August 2023; revised 1 December 2023; accepted 20 December 2023. Date of publication 17 January 2024; date of current version 9 April 2024. This work was supported in part by the UCLM-Telefónica Chair of "Advanced Interaction Systems for Digital Education" and the Research Projects CODIFICA of the Ministry of Science and Innovation, Spain, under Grant PID2021-125122OB-I00, and in part by the FRAWA of the Junta de Comunidades de Castilla-La Mancha (JCCM) and the European Regional Development Fund (FEDER) under Grant SBPLY/21/180501/000244. *(Corresponding author: Ana I. Molina.)*

Christian X. Navarro-Cota is with the Faculty of Engineering, Architecture and Design, Autonomous University of Baja California, Ensenada 22860, Mexico (e-mail: cnavarro@uabc.edu.mx).

Ana I. Molina, Miguel A. Redondo, and Carmen Lacave are with the Department of Computer Science, University of Castilla-La Mancha, 13071 Ciudad Real, Spain (e-mail: AnaIsabel.Molina@uclm.es; Miguel.Redondo@uclm.es; Carmen.Lacave@uclm.es).

Digital Object Identifier 10.1109/TE.2023.3347191

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

the scientific community, even considering specific elements aimed at ensuring usability [17], [18]. An important consequence emerging from research in this area [19] is that pedagogical aspects are not taken into account for usability evaluation [20]. In fact, the concept of *pedagogical usability* is proposed in [21], and [22], being defined as the ability of an educational system or tool to be used effectively and efficiently in teaching and learning processes, facilitating the achievement of educational objectives.

Although students are more likely to have a satisfactory learning experience when using a well-designed learning software application that presents acceptable levels of technical or pedagogical usability from various angles [23], the literature shows clear evidence that the balance is too heavily weighted toward evaluating the usability of the technical aspects, exclusively [1]. Consequently, there is a lack of instruments to adequately measure both dimensions: the technical and the pedagogical (see Section II). And the few that exist do not meet standard quality criteria in terms of *validity* and *reliability* of their results [24], [25], so they should not be used [26] since the generation of scientific knowledge with a desirable level of precision and certainty is not guaranteed [27].

Consequently, this *work aims* to create an evaluation tool for m-learning applications, properly validated, that considers both pedagogical aspects and usability requirements. This instrument has the potential for dual functionality: an evaluation tool and a checklist or set of heuristics to guide the design processes of m-learning systems. Based on this goal, the following research questions are formulated.

RQ1: How could an instrument be developed to evaluate the usability of m-learning applications that combine technical and pedagogical aspects?

RQ2: How can the quality of the developed instrument be determined?

To answer these questions, the process of constructing an evaluation instrument called CECAM (*Cuestionario de Evaluación de la Calidad de Aplicaciones M-learning*," for its initials in Spanish[1]) is described,as is how it has been calibrated with statistical methods to demonstrate its quality [28]. This instrument, unlike other heuristics and guides for designing and evaluating m-learning applications, considers both technological and pedagogical aspects.

The remainder of this article is organized as follows: Section II includes a review of related works in the field of m-learning evaluation; Section III presents the main characteristics of the mobile learning evaluation framework (MOLEF) framework, on which the proposed measurement instrument (CECAM) is based; Section IV describes the process of developing the CECAM questionnaire, while its quality analysis or validation is detailed in Section V. Finally, the Discussion (Section VI) and Conclusions (Section VII) sections are included.

II. Related Works

The importance that *mobile learning*, or m-learning, is currently acquiring is unquestionable, given the massive presence of mobile phones in all fields, including education [29]. Their use is growing in recent years, and even more, since the COVID-19 pandemic, which revealed the need for more flexible, real-time, and remote access to educational resources [30], [31].

The m-learning approach enables extended learning, exploring the ubiquitous possibilities of technologies, such as laptops, smartphones, or tablets—to access, record, process, manage, and exchange information anytime, anywhere.

However, the proper design of this type of applications continues to be a challenge [32], [33]. Considering usability aspects when creating mobile learning applications is essential to improve user acceptance and satisfaction, increasing their motivation and engagement. The most frequently reported usability aspects of m-learning in the literature have been [11]: learnability, user satisfaction, ease of use, and usefulness. Considering usability aspects will also result in the creation of more motivating, effective, and efficient learning experiences for learners [34].

The interest in the *usability of m-learning* is growing in recent years, and there are several systematic literature review that try to know the current state on this topic [11], [14]. These works conclude that further research is needed in this area since many of the existing proposals are based on usability methods and standards for nonmobile or noneducational applications. At the support device level, there are notable differences between mobiles and desktops, such as touch interaction, ubiquity, limited screen size, and a greater demand for visual attention, which affect usability and should be taken into account in the design and evaluation of this type of applications [35]. On the other hand, it is necessary to consider educational aspects, to ensure that the use of these applications helps students in their learning process, in various contexts of use, and accordance with the posed learning objectives [22]. Thus, the concept of *pedagogical usability* emerges, which takes into account the learning process, learning purposes, user needs, learning experience, learning content, and learning outcomes [21].

Several frameworks, taxonomies and guidelines have been proposed in recent years for the design of m-learning applications and the evaluation of their usability. Table I shows some of the most outstanding proposals. Some of these works are based on standards, such as ISO 9241 [36], ISO/IEC 14 598 [37], ISO/IEC 25 000 [38], and IOS/IEC 9126 [39]. Others are based on well-established and validated frameworks, such as the theory of reasoned action (TRA) [40] or the DeLone and McLean model of information systems success (DL&ML) [41]. But, undoubtedly, the framework that has been most often taken as a reference is the technology acceptance model (TAM) [42], [43], proposed by Davis [44]. Works based on this framework seek to identify which factors best explain user intentions, as well as the adoption and acceptance of m-learning solutions. Some of these proposals include instruments, in the form of checklists or questionnaires, which allow quantifying the usability of the evaluated m-learning system.

As can be seen in Comparative Table I, only three proposals consider pedagogical aspects, with only two specifying

[1]The name of the instrument translated into English: *Questionnaire to Evaluate the Usability of M-learning Applications.*

TABLE I
RECENT AND NOTEWORTHY PROPOSALS THAT ADDRESS THE EVALUATION OF M-LEARNING SYSTEMS

Authors, year Reference	Proposal or contribution. Based on…	Consider pedagogical aspects	Propose a quantitative measurement instrument	Validated instrument? Validated aspects of the instrument.	Proposal highlights
Althunibat et al., 2022 [45]	• 8 factors of user satisfaction (5 technical and 3 pedagogical). • Based on the IOS/IEC 9126 and DL&ML model.	**Yes**	Questionnaire (58 items, 5-point Likert scale) for measuring 8 factors.	• Reliability: Cronbach's α.	• **Quality model** for m-learning application for **children** (aged between 3 and 12 years). • Considers technical and **pedagogical** factors.
Qashou, 2021 [46]	• A conceptual model to examine the factors related to student's behavioral intention to use m-learning in higher education. • Based on the TAM framework.	No	Questionnaire (28 items, 5-point Likert scale) for measuring 7 factors.	• Reliability: Cronbach's α. • Convergent validity: composite reliability (CR) and average variance extracted (AVE). • Discriminant validity.	• Combines TAM factors with some external constructs (**mobility**, **enjoyment** and **self-efficacy**).
Limtrairut, 2020 [47]	• A set of 16 usability heuristics for m-learning applications. • Based on Nielsen's heuristics and a review of the literature.	**Yes**	No	No	• Considers heuristics specific to e-learning and mobile interfaces, to enrich the **pedagogical and mobility aspects** respectively.
Kumar and Goundar, 2019 [48] [49]	• A set of 13 usability heuristics for m-learning applications. • Based on Nielsen's heuristics and a review of the literature.	No	No	No	• Provides a **guide for conducting heuristic evaluations** of m-learning.
Almaiah and Alismaiel, 2019 [50]	• A model to investigate the effects of quality factors and individual beliefs on student satisfaction and intention to use. • Based on integration of TAM framework with DL&ML model and literature review.	No	Questionnaire (31 items, 5-point Likert scale) for measuring 8 factors.	• Reliability: Cronbach's α and the Exploratory Factor Analysis (EFA).	• It allows examining the effective **factors underlying technology acceptance and usage behavior**, to design more effective and higher quality m-learning systems.
Parsazadeh et al., 2018 [51]	• 8 usability attributes (effectiveness, efficiency, timeliness, satisfaction, learnability, memorability, error and cognitive load). • Based on Nielsen's heuristics and ISO 9241.	No	Questionnaire (42 items, 5-point Likert scale) for measuring 8 factors.	• Reliability: Cronbach's α. • Delphi method to validate the questionnaire with a group of 14 experts.	• Inclusion of the **timeliness** (interactive response time) attribute.
Ahmad, Rextin, and Kulsoom, 2018 [52]	• Set of 25 guidelines in 7 categories or factors. • Based on a literature review.	No	A brief questionnaire (8 items, 5-point Likert scale) for measuring one factor (satisfaction).	• Reliability of satisfaction questionnaire: Cronbach's α.	• Develop a comprehensive **list of usability guidelines** suitable for multiple platforms and genres of smartphone applications.
Buabeng-Andoh, 2018 [53]	• A model to predict and explain university students' intention to use m-learning. • Based on the integration of TAM and TRA frameworks.	No	Questionnaire (23 items, 7-point Likert scale) for measuring 5 factors.	• Convergent validity: 3 methods to examine this aspect: item reliability, composite reliability (CR), and average variance extracted (AVE). • Discriminant validity: Fornell-Larcker criterion.	• Explores the ability of the **integration of TAM and TRA** to predict and explain university students' intention to use m-learning in classrooms.
Soad, Duarte Filho and Barbosa, 2016 [54]	• A quality model, considering a set of 8 quality criteria, and a set of metrics or quality attributes. • Based on best practices and knowledge established by ISO/IEC 14598 and ISO/IEC 25000, and a systematic literature review.	**Yes**	A checklist, which comprises 80 questions, for measuring the 8 quality criteria.	No	• Proposes a **quality model** of m-learning applications. • Proposes a catalog of quality attributes divided **into technical, educational, sociocultural, and socioeconomic** characteristics.
Sarrab, Elbasir and Alnaeli, 2016 [55]	• A quality model, considering a set of 12 technical quality aspects of m-learning. • Based on the DL&ML model and literature review.	No	No	No	• Proposes a **quality model** that captures the most abstract and generic technical aspects of m-learning **service quality**.

instruments for their measurement. However, in one proposal, the instrument has yet to be validated, and in the other, its validation has been limited to the calculation of Cronbach's Alpha. Therefore, there is a need to propose a measurement instrument, or questionnaire, that considers aspects of mobile and pedagogical usability, suitably validated, and refined, that

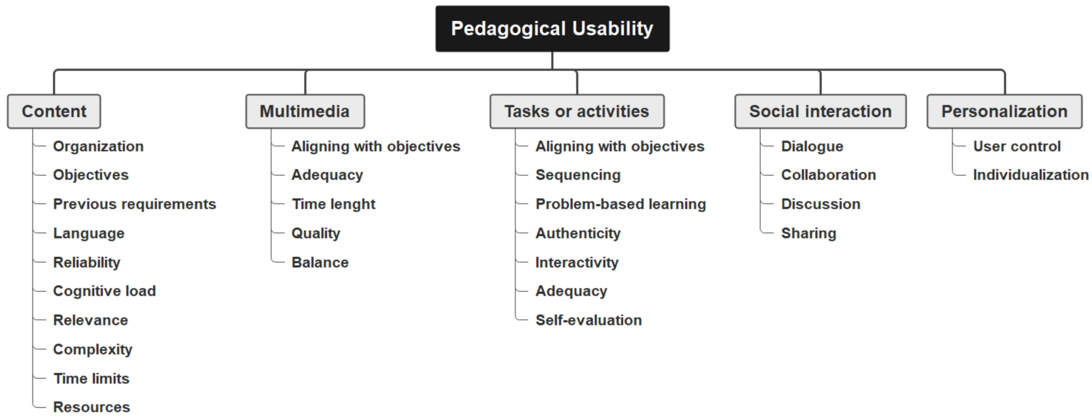

Fig. 1. Pedagogical Usability: Subdimensions and quality criteria.

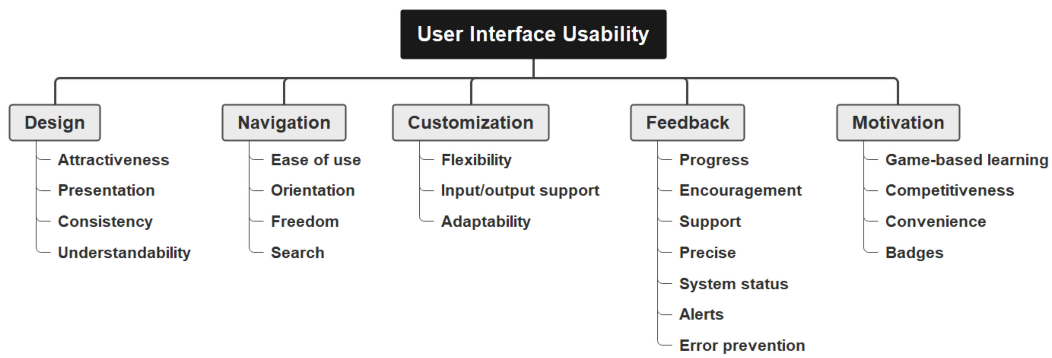

Fig. 2. User Interface Usability: Subdimensions and quality criteria.

will allow evaluators to determine the quality of an m-learning system [56].

Having reviewed the main existing proposals for the evaluation of m-learning systems, the following section briefly describes the proposed framework for evaluating these types of applications, called MOLEF, on which the measurement instrument created and described in this article is based.

III. MOLEF—Mobile Learning Evaluation Framework

As indicated in the Introduction section, this article presents the process of developing and validating the CECAM questionnaire, a measuring tool proposed to evaluate and design m-learning applications. It considers the elements in the framework MOLEF described in [33]. This framework was developed after a thorough analysis of the existing evaluation and development frameworks for m-learning applications, as well as of well-known and widespread models of technology adoption, such as TAM [44] or Unified Theory of Acceptance and Use of Technology (UTAUT) [57], among others. This analysis has led to the identification of a series of factors or quality requirements [58], which comprise the MOLEF framework.

The MOLEF framework considers pedagogical factors (e.g., aligning with learning objectives, adequacy, cognitive load), mobile usability features (e.g., adaptability, consistency, flexibility), as well as technology adoption factors (e.g., usefulness or relevance, ease of use, previous requirements).To this end, it proposes a catalog of usability attributes or characteristics of m-learning systems [58], based on a literature review on the design and evaluation of mobile applications, the quality standard ISO/IEC: 25010:2011 2011 [59], adoption factors, and pedagogical usability attributes.

The proposed set of quality attributes is divided into two main blocks or dimensions: 1) pedagogical usability (Fig. 1) and 2) user interface usability (Fig. 2).

Each of the main dimensions (pedagogical and user interface usability) includes a subset of subdimensions that, in turn, are divided into a set of criteria or quality attributes:

1) The *pedagogical usability* considers educational and pedagogical factors to support learning activities. These factors will provide the appropriate context for educational practice. This dimension establishes five subdimensions: a) *content*; b) *multimedia*; c) *tasks or activities*; d) *social interaction*; and e) *personalization*. Each subdimension includes a set of quality criteria (Fig. 1), the definition of which can be found in [58].
2) The *user interface usability* includes factors that make the software easier to use and that favor the acceptance and satisfaction of the students with the m-learning system. This dimension includes five subdimensions related to the interaction with the interface: a) *design*; b) *navigation*; c) *customization*; d) *feedback*, and

e) *motivation*. Each subdimension is measured by a set of quality criteria (Fig. 2) [58].

IV. Development of the CECAM Measurement Tool

This section presents the *process of developing the CECAM questionnaire (RQ1)*, a measuring tool proposed to evaluate and design m-learning applications. It considers the elements in the framework MOLEF [33] (see Figs. 1 and 2).

A. Questionnaire Design

Questionnaire surveys are helpful tools used to gather information or measure. For instance, questionnaires are commonly used as a measuring tool for educational software [60].

Therefore, a structured questionnaire was created as a tool for evaluating m-learning applications (CECAM). The questions require short, concrete, and closed answers (choosing between a range of five options), which should be prepared beforehand. Particularly, each question is an affirmation that describes an attribute that should be considered in m-learning applications. The evaluator will mark ≪X≫ to register the degree of fulfillment of such attributes, using the following scale: 1) *Strongly disagree*; 2) *Disagree*; 3) *Neither*; 4) *Agree*; and 5) *Stronglyagree*. These five options have been chosen because they cover all the possible answers (principle of exhaustively), avoiding the possibility of evaluators not responding due to a lack of available responses and preventing the possibility for the evaluator from choosing two answers for the same question (to guarantee the exclusivity of the questionnaire) [61].

B. Questionnaire Structure

Considering the categories in MoLEF a questionnaire was developed and divided into two multidimensional subscales (pedagogical usability and user interface usability). Therefore, the questions are grouped into dimensions with items related to the aspects and constructs mentioned in the framework. Specifically, the questionnaire has some questions or heuristics related to each established criterion. Therefore, the pedagogical usability scale includes questions related to the content, multimedia, activities, social interaction, and personalization; and the user interface usability scale contains the questions related to the interface design, navigation, customization, feedback and motivation.

Table II shows the preliminary structure of the CECAM questionnaire, which includes the subscales, the constructs, and the number of items of these constructs or factors to be measured.

Once the questionnaire structure has been presented, the process followed in elaborating the initial list of items that compose it is described.

C. Elaborating Initial Items

The items in the questionnaire were written based on the factors of MOLEF [33], focusing on presenting a clear and understandable language, avoiding the use of technical words that would prevent a clear interpretation by the evaluators, and therefore ensuring: 1) that the exact question is being answered without misinterpretations of the statement; 2) avoiding the possibility of questions being unanswered due to lack of comprehension; and 3) not making it too complicated for the evaluator.

TABLE II
Subscales, Constructs, and Number of Items of the Preliminary Questionnaire

CECAM QUESTIONNAIRE	SUBSCALES	CONSTRUCTS OR FACTORS
Evaluation of the Usability of m-Learning Applications (72 items)	Pedagogical Usability (33 items)	Content (9 items)
		Multimedia resources (7 items)
		Task or activities (9 items)
		Social interaction (4 items)
		Personalization (4 items)
	User Interface Usability (39 items)	Design (6 items)
		Navigation (13 items)
		Customization (8 items)
		Feedback (7 items)
		Motivation (5 items)

Table VI (in Appendix) presents a description of the items that belong to each of the constructs in the pedagogical usability subscale and their assigned identifier (ID) for further questionnaire statistics analysis. Table VII (in Appendix) presents a description of the items that belong to each of the constructs in the user interface usability subscale and an ID assigned for further statistical analysis of the questionnaire.

The first group of 72 items forms the initial structure of the preliminary questionnaire. This first version is the result of a revision performed by two experts in the area of evaluation questionnaires. The main goal was to review the wording of the items and analyze phrases that could confuse the evaluators. As a result of this revision, some items were modified; some questions changed from negative to positive; some terms were eliminated to avoid confusion; and examples or explanations in parentheses were included to facilitate understanding of each item's questions whenever necessary.

V. Quality Analysis and Results of the CECAM Survey

The previous section described the process followed to develop the CECAM questionnaire, which provides an affirmative answer to the research question *RQ1*, on whether it is possible to develop an instrument to evaluate the usability of m-learning applications, combining technical and pedagogical aspects. The answer to the *RQ2* research question, about *how way to measure the quality of the developed instrument*, is given by the calibration of the CECAM survey in terms of the standard criteria of quality: *reliability* and *validity* [24].

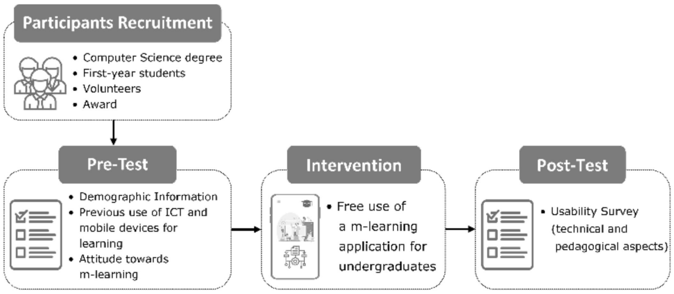

Fig. 3. Experimental design.

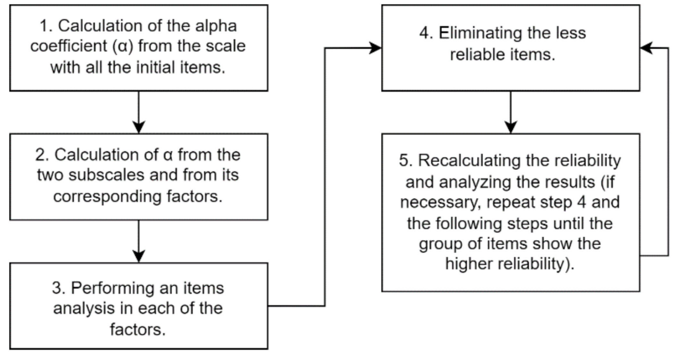

Fig. 4. Process to calculate the reliability for the CECAM questionnaire.

The validity of a survey should be interpreted as the degree to which evidence and theory support the interpretations of test scores; the reliability of a test provides the degree of consistency or stability of measures when the measurement process is repeated [28]. The data needed for the application of the method used to calibrate the questionnaire [25], have been obtained through a quasi-experimental study, which is described in the following section. All statistical procedures were performed with *Software IBM SPSS*, version 21.

A. Quasi-Experiment Design to Obtain Calibration Data

With the aim of obtaining the data needed for the calibration of the CECAM questionnaire, a quasi-experiment [62] was designed to be carried out with university students without random assignment. Fig. 3 illustrates the phases involved in the process: recruitment of participants, pre-testing, intervention, and post-testing, which are described below.

The participants were recruited from the first-year students of the computer science (CS) degree taught in the College of Computer Science (CCS) of the University of Castilla-La Mancha (UCLM), Spain. All enrolled students in the compulsory class of *Programming I*, taught in the first semester of the CS degree, were informed in class of the experience to be carried out (what it consisted of and the estimated time it would take) and that they would receive a reward in the form of some extra points for the final grade of the course. As a result, 37 students voluntarily decided to participate in the experience. The small sample size is a reflection of the usually low-attendanceat classes by students, which several reasons can explain: 1) class attendance is not compulsory; 2) all the material to follow the subject is available to them through the university's online platform (Moodle); and 3) *programming* is usually one of the most difficult subjects for CS students [63] which implies that many students drop it.

Then, the pre-test phase involved providing each participant with a paper copy of a survey to gather some personal information about them, and they must fill it out anonymously during the initial 20 min of the class. The first block of questions asked for demographic information, such as age, gender, and level of education. The second block included questions about their experience in using Information Communication Technology (ICT) and mobile devices for learning, and their attitude toward mobile learning. Then, the intervention consisted of using a m-learning application to learn a particular course topic for a specified period. The chosen application was *Learn Java-Free*,[2] available through Google Play. The application's goal is to review topics of Java programming, and for that purpose, it contains several concepts included in the subject contents. The students were suggested to use it for three weeks to practice the Java concepts explained during laboratory classes and were allowed to use it freely. Finally, during the post-testing, participants had to complete the CECAM survey to measure the application's usability.

B. Reliability Analysis

The analysis of the reliability of a survey involves calculating its *internal consistency,* which measures whether several items intended to measure the same general construct produce similar scores, and the *homogeneity index* of each item, which indicates the degree to which each item contributes to the internal consistency of the scale [25]. The internal consistency is usually obtained by *Cronbach's alpha* (α)coefficient [64], which is based on the average interitem correlation and assumes that the items (measured on the Likert scale) assess the same construct and are highly correlated. The values of this coefficient vary between 0 and 1, considering that the closer is to 1, the higher the internal consistency. A general rule considers a coefficient *acceptable* when its value is equal to or greater than 0.7 [65]. For multidimensional scales, each of one measuring different aspects, the calculation of the internal consistency is performed also on each dimension. In addition, it is advisable to evaluate the value of the αcoefficient after removing each of the items from the survey in turn: items whose removal make the coefficient value increase can be disregarded. Regarding the homogeneity index of each item, it is defined by the *Pearson correlation* coefficient between the scores of the item and the sum of the scores on the remaining items. Items with low homogeneity indices measure something other to what is reflected by the survey, so they can be removed.

It is usually advisable to remove those items whose homogeneity index is less than 0.35 [66]. In this case, the process is repeated with the remaining items until all have a homogeneity index greater than 0.35, as Fig. 4 shows.

[2]https://play.google.com/store/apps/details?id=com.rfb.learnjavafree&hl=en_US

TABLE III
RELIABILITY OF THE PEDAGOGICAL USABILITY SUBSCALE AND ITS FACTORS

Pedagogical Usability (α=0.892)														
Content (α=0.676)			**Multimedia resources (α=0.895)**			**Task or activities (α=0.866)**			**Social interaction (α=0.916)**			**Personalization (α=0.816)**		
Item	Item-total correlation	α without item	Item	Item-total correlation	α without item	Item	Item-total correlation	α without item	Item	Item-total correlation	α without item	Item	Item-total correlation	α without item
C1	0.523	0.625	M1	0.696	0.83	A1	0.659	0.846	S1	0.672	0.935	P1	0.644	0.765
C2	0.607	0.606	M2	0.685	0.833	A2	0.706	0.843	S2	0.906	0.857	P2	0.680	0.748
C3	0.406	0.64	M3	0.410	0.869	A3	0.701	0.844	S3	0.805	0.892	P3	0.678	0.75
C4	0.545	0.614	M4	0.762	0.819	A4	0.624	0.850	S4	0.869	0.871	P4	0.547	0.809
C5	**0.339**	**0.654**	M5	0.609	0.843	**A5**	**0.369**	**0.873**						
C6	0.42	0.634	M6	0.568	0.848	A6	0.657	0.849						
C7	**0.147**	**0.693**	M7	0.694	0.831	A7	0.608	0.853						
C8	**-0.022**	**0.735**				A8	0.65	0.849						
C9	0.447	0.627				A9	0.558	0.859						

TABLE IV
RELIABILITY OF THE USER INTERFACE USABILITY SUBSCALE AND ITS FACTORS

User Interface Usability (α=0.798)														
Design (α=0.817)			**Navigation (α=0.699)**			**Customization (α=0.673)**			**Feedback (α=0.489)**			**Motivation (α=0.556)**		
Item	Item-total correlation	α without item	Item	Item-total correlation	α without item	Item	Item-total correlation	α without item	Item	Item-total correlation	α without item	Item	Item-total correlation	α without item
D1	0.455	0.822	N1	0.349	0.679	CU1	0.385	0.848	**F1**	**-0.453**	**0.703**	**MO1**	**0.195**	**0.566**
D2	0.592	0.785	N2	0.619	0.641	CU2	0.498	0.846	F2	0.382	0.377	**MO2**	**0.355**	**0.478**
D3	0.809	0.739	N3	0.378	0.675	CU3	0.401	0.850	F3	0.490	0.317	**MO3**	**0.245**	**0.551**
D4	0.597	0.784	N4	0.440	0.664	**CU4**	**-0.080**	**0.864**	**F4**	**0.216**	**0.461**	**MO4**	**0.308**	**0.506**
D5	0.456	0.812	**N5**	**0.236**	**0.693**	CU5	0.609	0.858	**F5**	**0.193**	**0.467**	**MO5**	**0.518**	**0.539**
D6	0.622	0.778	N6	0.426	0.671	CU6	0.639	0.863	F6	0.424	0.354			
			N7	0.612	0.644	**CU7**	**0.228**	**0.854**	F7	0.577	0.521			
			N8	0.428	0.667	CU8	0.299	0.873	**F12**	**0.349**	**0.757**			
			N9	**-0.068**	**0.735**									
			N10	0.599	0.635									
			N11	**-0.041**	**0.727**									
			N12	**0.261**	**0.691**									
			N13	**0.065**	**0.715**									

In the case of the CECAM questionnaire, $\alpha = 0.9$, representing an excellent internal consistency. The pedagogical usability subscale has also a good internal consistency ($\alpha = 0.892$), and the user interface usability subscale has an acceptable internal consistency ($\alpha = 0.798$), as Tables III and IV, respectively, shows in their first row. Regarding the pedagogical usability subscale, Table III shows the α value of each of its dimensions and their factors, highlighting in bold the coefficients below 0.7. That of the Content factor is lower than 0.70 ($\alpha = 0.676$), but after removing items C5, C7 and C8 (see Table VI of Appendix) it grows to 0.735, which is acceptable. Moreover, the removing has sense because item C1 can substitute to item C5; item C4 can replace item C7, and item C8 does not explicitly evaluate a pedagogical quality. The other factors have a good internal consistency, although that of the Task or activities could improve if item A5 were removed. Despite its homogeneity index is greater than 0.2, it was removed since it was considered a difficult question to understand, and A4 and S1 can replace it. The final internal consistency of the reduced pedagogical subscale is excellent ($\alpha = 0.900$).

Concerning the factors of the user interface subscale, Table IV shows that only the Design-factor has a good internal consistency ($\alpha = 0.817$); that of the Navigation-factor is close to be acceptable ($\alpha = 0.699$) and it grows ($\alpha = 0.799$) after removing items N9, N11, N12 and N13 (see Table VII of Appendix): N8 could replace N9, and N10 can replace N11, N12 can be considered a feature that depends more on the network connection than the application, and N13 was a suggested but not essential feature. The case of the Customization-factor is similar because its consistency

TABLE V
SUBSCALES, CONSTRUCTS, AND NUMBER OF ITEMS OF THE FINAL QUESTIONNAIRE

CECAM QUESTIONNAIRE	SUBSCALES	CONSTRUCTS OR FACTORS
Evaluation of the Usability of m-Learning Applications (56 items)	Pedagogical Usability (29 items)	Content (6 items)
		Multimedia resources (7 items)
		Task or activities (8 items)
		Social interaction (4 items)
		Personalization (4 items)
	User Interface Usability (27 items)	Design (6 items)
		Navigation (9 items)
		Customization (6 items)
		Feedback (6 items)

TABLE VI
ID and Description of Each Item in the Pedagogical Usability Subscale

ID	Item
Pedagogical Usability	
Content	
C1	The content is organized in modules or units.
C2	The objectives are defined at the beginning of the module or unit.
C3	The students are informed about previous abilities or knowledge required, if necessary.
C4	The explanation of the concepts is presented in a clear and concise way.
C5	The time invested on each module or unit lasts less than 7 minutes.
C6	The modules or units are organized according to the level of difficulty (from easier to more complex).
C7	The language is appropriate for the students.
C8	The content is free of errors.
C9	There are links to external resources related to the contents and adapted to mobile devices.
Multimedia resources	
M1	Different types of multimedia resources are presented (videos, images, audios, animations, simulations, etc.), related to the learning objectives.
M2	Multimedia resources have been selected appropriately to facilitate learning.
M3	Multimedia resources length is less than 7 minutes.
M4	The multimedia contents have a good quality for video, audio and images.
M5	The multimedia resources can be downloaded to the mobile device.
M6	The multimedia resources have the appropriate size to be downloaded to a mobile device.
M7	There is an appropriate proportion of multimedia resources.
Tasks or activities	
A1	Activities are proposed to acquire new skills that determine their learning (e.g., questions, associations, exercises, problem solving, etc.).
A2	The activities facilitate the comprehension of the educational contents.
A3	The learning activities help improving and strengthen abilities.
A4	Activities allow students integrate new information with previous learning.
A5	Activities promote creativity, by allowing the students generate their own solutions.
A6	Activities reflect practice relevant to real or professional life.
A7	The activities are congruent to the student's competences (not to easy, and not to complex).
A8	There are activities to assess the learning of educational content (e.g., tests, assessments, exercises etc.).
A9	The activities take benefit of the functions and advantages of the use of mobile devices (photography, recording videos or audios, augmented reality, simulations, QR codes, etc.).
Social interaction	
S1	There are opportunities for students to perform group projects.
S2	The application allows communication with classmates or teachers to solve doubts about the contents (chat, email, etc.).
S3	The application allows sharing information like photos, videos or documents related to the activities, in order to discuss them (ex. through social media, blogs, wiki, etc.).
S4	The system allows opportunities for competition between students (ex. Visualizing the achievements of other students in the group).
Personalization	
P1	The application allows the student to create their own learning route.
P2	The application allows the evaluation of the students' knowledge and suggests contents to study according to results.
P3	The application allows the student to choose among different levels of complexity.
P4	The application allows the student to establish study goals (ex. define minutes daily or weekly).

is near acceptable ($\alpha = 0.673$) and it grows ($\alpha = 0.741$) after removing items C4 and C7: C4 was considered unnecessary, and C7 can be replaced by the feature measured by C6. In the Feedback-factor, the consistency is poor ($\alpha = 0.489$). Moreover, Motivation-factor has not enough consistency ($\alpha = 0.556$) and it does not grow after removing any item. Concerning the Motivation factor, a detailed analysis of their items revealed that they should be included in the Feedback factor. Therefore, items M1 to M5 (see Table VII of Appendix) were moved to that factor and renamed as F8 to F12, respectively. The Cronbach's alpha for the new Feedback increased ($\alpha = 0.756$) after removing items F1, F4, F8, F10, and F11. Finally, the alpha in the subscale stayed at 0.798.

C. Validity Analysis

The study of the validity of a teaching survey involves the following analysis [25].

1) *Content Validity,* which assesses the understanding of statements, is typically determined through expert judgment, where qualified individuals evaluate the survey [67]. The content of the CECAM questionnaire has been validated by 10 experts in the usability field, from the UCLM, who acted as judges. They all were given the same document that clearly stated the purpose of the survey, its content (Table I) and a rubric, Table II, specifying how they should make their evaluation of the defined dimensions, the items associated with each, and the evaluation scale. Given that 9 of 10 of the judges agreed to keep the original 10 dimensions and 72 items, as well as the Likert-type scale, the content of the questionnaire remained unchanged from the original proposal.
2) *Construct Validity,* which evaluates the degree to which an instrument reflects the theory of the concept that measures [68]. There are different methods for this type of analysis, although convergent and discriminant validity are commonly used [69].
   a) *Convergent validity* verifies that the items in the scales are significantly and strongly correlated with the constructs they belong to [68]. Among the different criteria to analyze the convergent validity,

TABLE VII
ID AND DESCRIPTION OF EACH ITEM OF THE USER INTERFACE USABILITY SUBSCALE

ID	Item
User interface usability	
Design	
D1	The interface design is aesthetic and attractive to the student.
D2	The font type, size and spaces allow easy reading of the information.
D3	The design has the appropriate number of colors, and does not generate visual overloading (2 to 4 maximum).
D4	The colors contrast is appropriate (ex. dark text over a clear background).
D5	The information presented adjusts or adapts to the size of the screen.
D6	The interface design is consistent in style, font size, buttons, colors, etc. (same design through the whole application).
Navigation	
N1	The main menu and main options have an appropriate visibility.
N2	The navigation is simple, familiar and intuitive (ex. it is easily understood what has to be done).
N3	The desired contents or basic tasks are easily accessible from the main page in three or less clicks.
N4	The application informs the part of the process where the student is (ex. title of the unit of learning selected, contrast of color on the options selected, etc.).
N5	The dimension and proximity of the tactile buttons or selection controls are appropriate to be easily selected with finger.
N6	The placement of the tactile buttons or others are reachable using mainly only one hand.
N7	Icons or elements represent actions and are familiar and intuitive (it is easy to recognize what it represents).
N8	The application clearly presents the option that allows to go back to the main menu.
N9	The application allows going back or exiting a non-desired screen, when it has been selected by error.
N10	The application clearly presents the search option to help students find contents.
N11	There are instructions of use for certain functions in the application, these are easy to find and understand (help option).
N12	The time between the selection of a function and the execution is relatively short.
N13	There is contextual help that guides the user during the use of the different interactive elements of the application.
Customization	
CU1	The application allows changing the font size and type.
CU2	The application allows changing the screen background and font color (ex. clear text over dark background and vice versa).
CU3	The application provides options of advanced configuration and are easy to find.
CU4	The application can be executed in different platforms (Android, iOS, Windows phone, etc.).
CU5	The application allows choosing among different languages.
CU6	The application allows you to choose different ways of input and output
CU7	The application allows you to convert text to speech, or transcribes text from a voice dictation.
CU8	The application allows basic functions over the contents (ex. underline or highlight text, making notes, copy and paste, etc.).
Feedback	
F1	The application provides precise and instant feedback related to the actions performed by the user.
F2	The application provides precise feedback related to the system state (ex. a status bar) showing information about the progress on an action).
F3	The application provides information about the units you already master and what you have yet to complete or study.
F4	The application provides reminders related to the goals stablished by the student.
F5	During the self-evaluations, if the student makes a mistake, the app will give the student opportunities to correct.
F6	If the student makes a mistake during the self-evaluations, the app will provide explanations about the correct responses.
F7	The application provides information when actions are performed that could generate non desired effects (ex. warnings or confirmations to prevent mistakes).
Motivation	
MO1	Students receive constructive or motivational feedback, when they have achieved a significant advance.
MO2	The application allows the accumulation of points after participating or completing activities.
MO3	The application defines a series of levels that must be achieved, in order to advance to the next level.
MO4	The application allows the student to obtain rewards when an objective is reached or a study goal is achieved. (ex. points, personalizing avatars, insignias, access to additional contents, etc.).
MO5	The application provides options for sharing significant progress or achievements (e.g., at the end of a level or at the end of a course) on social networks.

the factor loading matrix was chosen [70], since it calculates the Pearson correlation coefficient among the items and their constructs. It is recommended that this value should be higher than 0.4 [71]. This property is satisfied by all items of the pedagogical usability subscale, and those of the user interface usability subscale except for item N5 (see Table VII of Appendix). Nevertheless, it was decided not to remove it because it is an important characteristic of the usability.

b) *Discriminant Validity* is applied in the case of multidimensional scales, and tests if the different constructs that form it measure different concepts. Then, each item must be related to its construct and significantly different from the rest of the constructs to the ones it belongs [68]. Recommended methods to analyze the discriminant validity are the comparison between indicator correlations, and the comparison between the shared and the extracted variance. The comparison between the shared and the extracted variance is made through the analysis of the factor cross-loading matrix. It represents the Pearson correlation coefficients of the items and the other constructs. Discriminant validity exists if all correlations between the items of a construct are significant and each of these correlations is higher than all correlations between indicators of the other constructs. Table VIII (in Appendix) shows the cross-loading matrix for the factors, which reveals that each item is strongly associated with a particular construct, as evidenced by its high-factorial load

TABLE VIII
CORRELATION COEFFICIENTS OF THE ITEMS AND CONSTRUCTS OF THE SCALE (FACTORS CROSS-LOADING MATRIX)

Factors cross-loading matrix									
ID	Pedagogical Usability Subscale					User Interface Usability Subscale			
	Content	Multimedia resources	Task or activities	Social interaction	Personalization	Design	Navegation	Customization	Feedback
C1	**0.637**	0.038	0.339	0.131	-0.045	0.191	0.240	0.283	0.114
C2	**0.779**	0.266	0.377	0.260	0.235	0.186	0.382	0.263	0.449
C3	**0.626**	-0.017	0.494	0.294	0.424	-0.071	0.127	0.254	0.436
C4	**0.699**	-0.056	0.357	0.141	0.053	0.119	0.192	0.174	0.339
C6	**0.665**	0.084	0.283	0.423	0.236	0.154	0.213	0.437	0.266
C9	**0.714**	0.159	0.396	0.514	0.249	0.006	0.287	0.359	0.583
M1	0.147	**0.829**	0.407	0.403	0.152	0.437	0.420	0.309	0.128
M2	0.069	**0.814**	0.060	0.181	-0.094	0.327	0.497	-0.004	-0.183
M3	0.263	**0.544**	0.136	0.175	-0.009	0.324	0.403	0.156	0.036
M4	0.052	**0.866**	0.134	0.188	0.009	0.590	0.578	0.032	-0.073
M5	-0.018	**0.679**	0.340	0.263	0.269	0.196	0.304	0.064	-0.135
M6	-0.078	**0.614**	0.088	0.121	-0.028	0.294	0.184	-0.243	-0.150
M7	0.187	**0.804**	0.142	0.140	-0.042	0.479	0.414	0.035	-0.011
A1	0.595	0.279	**0.795**	0.370	0.259	0.109	0.448	0.337	0.272
A2	0.427	0.087	**0.793**	0.223	0.240	0.171	0.325	0.276	0.136
A3	0.208	0.040	**0.743**	0.268	0.122	0.067	0.254	0.119	-0.042
A4	0.227	0.231	**0.624**	0.364	0.312	-0.086	0.211	-0.022	-0.056
A6	0.438	0.099	**0.748**	0.446	0.529	0.094	0.313	0.330	0.116
A7	0.339	0.174	**0.681**	0.230	0.353	0.163	0.381	0.158	0.213
A8	0.507	0.310	**0.811**	0.562	0.441	0.373	0.494	0.455	0.196
A9	0.294	0.204	**0.682**	0.610	0.537	0.195	0.107	0.377	0.194
S1	0.427	0.282	0.518	**0.814**	0.477	0.136	0.312	0.327	0.270
S2	0.371	0.254	0.508	**0.946**	0.529	0.042	0.186	0.319	0.237
S3	0.385	0.224	0.477	**0.880**	0.504	0.052	0.163	0.333	0.113
S4	0.477	0.287	0.483	**0.933**	0.675	0.053	0.227	0.442	0.262
P1	0.101	-0.094	0.193	0.260	**0.704**	-0.132	0.028	0.067	-0.338
P2	0.207	-0.064	0.477	0.560	**0.852**	-0.119	-0.003	0.441	0.056
P3	0.258	0.024	0.274	0.344	**0.773**	0.096	0.246	0.044	0.000
P4	0.297	0.201	0.494	0.609	**0.827**	-0.013	0.188	0.307	0.229
D1	0.090	0.614	0.343	0.244	-0.040	**0.719**	0.417	0.160	0.044
D2	-0.072	0.361	-0.035	0.022	-0.105	**0.699**	0.191	-0.040	-0.209
D3	0.100	0.363	0.055	0.066	-0.037	**0.864**	0.282	0.059	-0.084
D4	0.224	0.270	0.187	-0.044	-0.121	**0.726**	0.391	-0.072	-0.023
D5	0.019	0.271	-0.086	-0.237	-0.098	**0.568**	0.421	-0.065	0.034
D6	0.107	0.245	0.252	0.106	0.152	**0.743**	0.336	0.235	0.175
N1	-0.097	0.488	0.049	0.101	-0.172	0.350	**0.492**	0.008	-0.259
N2	-0.051	0.377	0.034	-0.065	-0.252	0.349	**0.663**	-0.090	-0.180
N3	0.209	0.310	0.159	0.082	-0.002	0.504	**0.575**	-0.131	0.003
N4	0.367	0.530	0.520	0.307	0.175	0.291	**0.736**	-0.043	0.065
N5	-0.105	0.263	0.020	-0.065	0.010	0.298	**0.355**	0.171	0.004
N6	0.255	0.424	0.081	0.214	0.307	0.191	**0.604**	0.101	0.165
N7	0.343	0.395	0.338	0.327	0.290	0.140	**0.741**	0.123	0.042
N8	0.453	0.130	0.517	0.240	0.265	0.191	**0.622**	-0.004	0.032
N10	0.165	0.398	0.322	0.039	-0.092	0.522	**0.702**	-0.070	-0.053
CU1	0.019	-0.064	0.209	0.306	0.203	-0.116	-0.226	**0.482**	0.003
CU2	0.383	0.215	0.392	0.513	0.328	0.114	0.083	**0.822**	0.456
CU3	0.424	-0.003	0.283	0.010	0.204	-0.032	-0.046	**0.524**	0.253
CU5	0.244	0.026	0.155	0.208	0.074	-0.070	0.023	**0.807**	0.251
CU6	0.323	0.138	0.276	0.252	0.241	0.224	-0.019	**0.813**	0.264
CU8	0.209	-0.105	0.150	0.163	0.220	0.076	0.001	**0.453**	0.200
F2	0.358	-0.069	0.114	0.210	0.105	0.128	0.036	0.353	**0.712**
F3	0.575	0.123	0.327	0.339	0.084	0.021	0.143	0.257	**0.776**
F6	0.410	-0.278	0.053	0.156	0.002	-0.083	-0.258	0.399	**0.682**
F7	0.309	-0.167	-0.033	0.134	0.023	-0.039	0.036	0.198	**0.756**
F9	0.232	-0.008	0.147	-0.080	0.109	-0.111	-0.121	0.275	**0.581**
F12	0.256	0.296	0.135	0.056	-0.088	0.097	0.168	0.083	**0.443**

(indicated in bold). On the other hand, the comparison between the shared variance and extracted variance suggests that each construct should share more variance with its items than with the constructs in the scale. Therefore, the correlation coefficients between the constructs and the square roots of the average variance extracted (AVE) were calculated. For good discriminant validity, it is recommended that the square root of AVE in each construct be significantly higher than the correlations in the other constructs [72]. Table IX (in Appendix) shows the values of the square roots of the AVE for each construct (values in bold in the diagonal of the matrix) and the correlation

TABLE IX
CORRELATION COEFFICIENTS BETWEEN THE SQUARE ROOTS OF THE AVE OF EACH CONSTRUCT

Discriminant Validity Fornell-Larcker Criterion (*)									
	Pedagogical Usability Subscale					User Interface Usability Subscale			
	Content	**Multimedia resources**	**Task or activities**	**Social interaction**	**Personalization**	**Design**	**Navegation**	**Customization**	**Feedback**
Content	**0.689**								
Multimedia resources	0.137	**0.745**							
Tasks or activities	0.550	0.262	**0.737**						
Social interaction	0.468	0.295	0.555	**0.895**					
Personalization	0.320	0.053	0.506	0.617	**0.791**				
Design	0.123	0.531	0.220	0.080	-0.051	**0.725**			
Navigation	0.356	0.563	0.449	0.251	0.144	0.489	**0.621**		
Customization	0.440	0.107	0.395	0.401	0.336	0.090	-0.006	**0.671**	
Feedback	0.573	-0.050	0.208	0.251	0.074	0.006	-0.005	0.412	**0.668**

(*) For the discriminant validity to be met, the values of the diagonal (square root of the AVE of a construct) must be greater than the values below the diagonal (correlations between constructs).

coefficients with the other constructs (values under the diagonal). Therefore, all the construct values in the scale (see Table IX of Appendix), highlighted in bold, meet the requirements for discriminant validity.

Therefore, the quality analysis of the initial version of the CECAM questionnaire (Tables VI and VII of Appendix) has led to a reduced instrument (Table X of Appendix) with better reliability and validity. Nevertheless, the next section discusses the main findings obtained during the CECAM calibration.

VI. DISCUSSION

As can be derived from the study of related works (Section II), although in recent years the number of works proposing instruments to measure the usability of m-learning of a quantitative nature has grown, few of them consider the measurement of pedagogical aspects. Moreover, those that do so have not validated the quality of the instrument or have limited it to calculating Cronbach's Alpha. Based on this motivation, and in order to provide a solution to the detected need, in this work a survey instrument (CECAM) has been created, which includes the assessment of technological and pedagogical usability aspects (*RQ1*), and which has been conveniently validated and refined (*RQ2*). As a result of the reliability and validity analysis on the preliminary questionnaire CECAM (formed by 72 items and an initial alpha of 0.900), a scale with 56 items was obtained and an alpha reliability coefficient of 0.911. Based on the criteria by [65], the final version of CECAM questionnaire can be considered as an excellent measuring scale. Table V shows the final structure of the CECAM questionnaire, which contains 56 items.

Of the two subscales, one is considered excellent –pedagogical usability (0.900)– and the other acceptable –user interface usability (0.798). From the resulting constructs, four of them are considered acceptable –Content (0.769), Navigation (0.799), Customization (0.741), and Feedback (0.756); the other three as good –Multimedia resources (0.859), Educational activities (0.873) and Design (0.817) – and one as excellent –Social interaction (0.916). The 56-item CECAM questionnaire, described in Table X (in Appendix A), offers higher quality than the original questionnaire, both at the general level and for each of its specific dimensions and subdimensions. This suggests that, for further statistical analysis, the data provided by these 56 questions are more valid and more reliable than those provided by the 72 questions of the original questionnaire.

In the process of refining the instrument there is a subdimension that has been eliminated: the motivation dimension. This subdimension was part of the interface usability dimension, and considered aspects that increase learner engagement and motivation, related to gamification support [73], [74]. However, the aspects considered in this subdimension have shown not to be aligned with the rest of the subdimensions considered by the "classic" usability (usually Nielsen-based) and more commonly considered, such as navigation, feedback, etc. [75]. It is therefore proposed, as future work, to include a separate dimension in which aspects more related to the user experience (UX) would be considered [76]. UX is a more generic concept than classic usability and includes it, and in which subdimensions, such as value, desirability, engagement, entertainment, etc., can have place [77].

The instrument designed and validated was conceived as an evaluation support instrument and, therefore, as a tool for supporting usability evaluation methods (UEM) of *inquiry type*. Its use, combined with *testing methods*, allows to obtain quantitative measures of the quality (usability) of the m-learning system. However, the dimensions, subdimensions, and criteria included in the questionnaire can be considered as heuristics and design guidelines, which can be used both in usability *inspection methods* and in the design stages of mobile learning applications [78]. It is therefore considered that this questionnaire could be used in the initial (as a checklist or heuristic guide) and final (as an evaluation instrument) phases of the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) method [79].

In any case, and in relation to its initial objective, as a validation instrument, it is necessary to point out that in the area of Human-Computer Interaction and, specifically, in the field of UEM, it is proposed that the ideal is the combination of methods that provide more complete and comprehensive information on the usability of the interactive system than the independent and isolated application of a single evaluation method [80].

TABLE X
ITEMS INCLUDED IN THE FINAL VERSION OF THE CECAM QUESTIONNAIRE

ID	Item
Pedagogical Usability	
Content	
C1	The content is organized in modules or units.
C2	The objectives are defined at the beginning of the module or unit.
C3	The students are informed about previous abilities or knowledge required, if necessary.
C4	The explanation of the concepts is presented in a clear and concise way.
C5	The modules or units are organized according to the level of difficulty (from easier to more complex).
C6	There are links to external resources related to the contents and adapted to mobile devices.
Multimedia resources	
M1	Different types of multimedia resources are presented (videos, images, audios, animations, simulations, etc.), related to the learning objectives.
M2	Multimedia resources have been selected appropriately to facilitate learning.
M3	Multimedia resources length is less than 7 minutes.
M4	The multimedia contents have a good quality for video, audio and images.
M5	The multimedia resources can be downloaded to the mobile device.
M6	The multimedia resources have the appropriate size to be downloaded to a mobile device.
M7	There is an appropriate proportion of multimedia resources.
Tasks or activities	
A1	Activities are proposed to acquire new skills that determine their learning (e.g., questions, associations, exercises, problem solving, etc.).
A2	The activities facilitate the comprehension of the educational contents.
A3	The learning activities help improving and strengthen abilities.
A4	Activities allow students integrate new information with previous learning.
A5	Activities reflect practice relevant to real or professional life.
A6	The activities are congruent to the student's competences (not to easy, and not to complex).
A7	There are activities to assess the learning of educational content (e.g., tests, assessments, exercises etc.).
A8	The activities take benefit of the functions and advantages of the use of mobile devices (photography, recording videos or audios, augmented reality, simulations, QR codes, etc.).
Social interaction	
S1	There are opportunities for students to perform group projects.
S2	The application allows communication with classmates or teachers to solve doubts about the contents (chat, email, etc.).
S3	The application allows sharing information like photos, videos or documents related to the activities, in order to discuss them (ex. through social media, blogs, wiki, etc.).
S4	The system allows opportunities for competition between students (ex. Visualizing the achievements of other students in the group).
Personalization	
P1	The application allows the student to create their own learning route.
P2	The application allows the evaluation of the students' knowledge and suggests contents to study according to results.
P3	The application allows the student to choose among different levels of complexity.
P4	The application allows the student to establish study goals (ex. define minutes daily or weekly).
User interface usability	
Design	
D1	The interface design is aesthetic and attractive to the student.
D2	The font type, size and spaces allow easy reading of the information.
D3	The design has the appropriate number of colors, and does not generate visual overloading (2 to 4 maximum).
D4	The colors contrast is appropriate (ex. dark text over a clear background).
D5	The information presented adjusts or adapts to the size of the screen.
D6	The interface design is consistent in style, font size, buttons, colors, etc. (same design through the whole application).
Navigation	
N1	The main menu and main options have an appropriate visibility.
N2	The navigation is simple, familiar and intuitive (ex. it is easily understood what has to be done).
N3	The desired contents or basic tasks are easily accessible from the main page in three or less clicks.
N4	The application informs the part of the process where the student is (ex. title of the unit of learning selected, contrast of color on the options selected, etc.).
N5	The dimension and proximity of the tactile buttons or selection controls are appropriate to be easily selected with finger.
N6	The placement of the tactile buttons or others are reachable using mainly only one hand.
N7	Icons or elements represent actions and are familiar and intuitive (it is easy to recognize what it represents).
N8	The application clearly presents the option that allows to go back to the main menu.
N9	The application clearly presents the search option to help students find contents.
Customization	
CU1	The application allows changing the font size and type.
CU2	The application allows changing the screen background and font color (ex. clear text over dark background and vice versa).
CU3	The application provides options of advanced configuration and are easy to find.
CU4	The application allows choosing among different languages.
CU5	The application allows you to choose different ways of input and output
CU6	The application allows basic functions over the contents (ex. underline or highlight text, making notes, copy and paste, etc.).
Feedback	
F1	The application provides precise feedback related to the system state (ex. a status bar) showing information about the progress on an action).
F2	The application provides information about the units you already master and what you have yet to complete or study.
F3	If the student makes a mistake during the self-evaluations, the app will provide explanations about the correct responses.
F4	The application provides information when actions are performed that could generate non desired effects (ex. warnings or confirmations to prevent mistakes).
F5	The application allows the accumulation of points after participating or completing activities.
F6	The application provides options for sharing significant progress or achievements (e.g., at the end of a level or at the end of a course) on social networks.

VII. CONCLUSION, LIMITATIONS, AND FUTURE WORK

This article proposes an evaluation tool for educational applications with mobile support, which considers pedagogical aspects and usability requirements, called CECAM. The CECAM survey can have a dual use, as an evaluation tool, but also as a checklist or set of heuristics in m-learning system design processes. Its application improves the usability of mobile learning applications, as well as the learning experience of students. This proposal is a relevant contribution to the field of m-learning, since there are hardly any evaluation instruments that contemplate technological and pedagogical aspects, which have been properly validated.Therefore, it is considered that its use can help improve the quality of m-learning systems. Even so, there is still work to be done in this area, particularly in ensuring that mobile learning support tools are accessible, effective, and satisfactory for users, providing them with good learning experiences, and consequently, promoting educational success.

The survey instrument developed, consisting of 56 items grouped into two dimensions (pedagogical usability and usability of the user interface), presents high levels of validity and reliability. One of the main *limitations* of the work described is related to small sample size (37 students) used to calibrate the CECAM questionnaire, since it does not reach the 50 participants recommended for robust statistical analysis [81]. The low number of participants in the experience poses a threat to the *external validity* of the proposed instrument, which implies that the results obtained by using it should be considered with caution, although they are not necessarily erroneous. Furthermore, there is an inherent threat to the *internal validity* of experimental designs in which the subjective perception of learners is measured: researcher and participant bias. The use of subjective (perception-based) surveys has the drawback that responses may be biased, as participants sometimes say what they think the researcher wants to hear. This threat can be reduced by promising subjects that their responses will be treated anonymously, but this does not guarantee that their answers will be completely truthful and objective. Despite these limitations, the instrument designed is very useful for researchers and practitioners to improve the effectiveness of m-learning applications and seeks to improve their design and implementation.

Therefore, this work highlights the importance of analyzing the validity and reliability of any measurement instrument before using it to draw conclusions about the data collected. As *future work,* it is contemplated to include an additional dimension in which specific factors related to the UX, or as it is called in the e-learning field, the *learner experience*, will be included. Once this has been done, it is planned to replicate the validation work described with a larger sample of students from several universities in order to increase the precision of the psychometric properties of the proposed survey.

APPENDIX

See Appendix Tables VI–X.

ACKNOWLEDGMENT

The authors would like to thank all the students who voluntarily agreed to participate in the evaluation process described in this article.

REFERENCES

[1] H. Crompton and D. Burke, "The use of mobile learning in higher education: A systematic review," *Comput. Educ.*, vol. 123, pp. 53–64, Apr. 2018, doi: 10.1016/j.compedu.2018.04.007.

[2] M. A. Job, "Enhancing learning process by implementing M-learning paradigm in open university education," *Int. J. Appl. Eng. Res.*, vol. 12, no. 21, pp. 11082–11087, 2017.

[3] H. Crompton, "A historical overview of mobile learning: Toward learner-centered education," in *Handbook of Mobile Learning*, New York, NY, USA: Routledge, 2013, pp. 3–14.

[4] M. M. Grant, "Difficulties in defining mobile learning: Analysis, design characteristics, and implications," *Educ. Technol. Res. Dev.*, vol. 67, no. 2, pp. 361–388, Apr. 2019, doi: 10.1007/s11423-018-09641-4.

[5] M. Sharples, J. Taylor, and G. Vavoula, "A theory of learning for the mobile age: Learning through conversation and exploration across contexts," in *Medienbildung in neuen Kulturräumen*, Wiesbaden, Germany: VS Verlag für Sozialwissenschaften, 2010, pp. 87–99.

[6] Y. Mehdipour and H. Zerehkafi, "Mobile learning for education: Benefits and challenges," *Int. J. Comput. Eng. Res.*, vol. 3, no. 6, pp. 93–101, 2013. [Online]. Available: http://www.ijceronline.com/papers/Vol3_issue6/part3/P03630930100.pdf

[7] R. Christensen and G. Knezek, "Readiness for integrating mobile learning in the classroom: Challenges, preferences and possibilities," *Comput. Human Behav.*, vol. 76, pp. 112–121, Nov. 2017, doi: 10.1016/j.chb.2017.07.014.

[8] S. Criollo-C, S. Luján-Mora, and A. Jaramillo-Alcazar, "Advantages and disadvantages of m-learning in current education," in *Proc. IEEE World Eng. Educ. Conf.*, 2018, pp. 1–6, doi: 10.1109/EDUNINE.2018.8450979.

[9] F. Zahra, A. Hussain, and H. Mohd, "Usability evaluation of mobile applications; Where do we stand?," *AIP Conf. Proc.*, vol. 1891, no. 1, 2017, Art. no. 20056, doi: 10.1063/1.5005389.

[10] A. Dirin and M. Nieminen, "User experience evolution of M-learning applications," *Proc. Int. Conf. Comput. Support. Educ.*, 2017, pp. 154–161, doi: 10.5220/0006370301540161.

[11] P. Weichbroth, "Usability of mobile applications: A systematic literature study," *IEEE Access*, vol. 8, pp. 55563–55577, 2020, doi: 10.1109/ACCESS.2020.2981892.

[12] J. Nielsen, *Usability Engineering*. Cambridge, MA, USA: Morgan Kaufmann, 1994.

[13] A. Dix, J. Finlay, G. D. Abowd, R. Beale, *Human-Computer Interaction*. Hoboken, NJ, USA: Prentice Hall, 2009.

[14] B. A. Kumar and P. Mohite,"Usability of mobile learning applications: A systematic literature review," *J. Comput. Educ.*, vol. 5, no. 1, pp. 1–17, 2018, doi: 10.1007/s40692-017-0093-6.

[15] L. Findlater and J. McGrenere, "Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces," in *Proc. Conf. Human Factors Comput. Syst.*, 2008, pp. 1247–1256.

[16] M. Sharples, J. Taylor, G. Vavoula, R. Andrews, and C. Haythornthwaite, "The SAGE handbook of e-learning research," in *A Theory of Learning for the Mobile Age*. Thousand Oaks, CA, USA: Sage Publ., 2007, pp. 221–247.

[17] J. Lu, M. Schmidt, M. Lee, and R. Huang, "Usability research in educational technology: A state-of-the-art systematic review," *Educ. Technol. Res. Dev.*, vol. 70, no. 6, pp. 1951–1992, 2022.

[18] A. I. Gunesekera, Y. Bao, and M. Kibelloh, "The role of usability on e-learning user interactions and satisfaction: A literature review," *J. Syst. Inf. Technol.*, vol. 21, no. 3, pp. 368–394, 2019.

[19] C. X. Navarro, A. I. Molina, M. A. Redondo, and R. Juárez-Ramírez, "Framework to evaluate M-learning systems: A technological and pedagogical approach," *Rev. Iberoam. Tecnol. del Aprendiz.*, vol. 11, no. 1, pp. 33–40, Feb. 2016, doi: 10.1109/RITA.2016.2518459.

[20] A. Jalil, M. Beer, and P. Crowther, "Pedagogical requirements for mobile learning: A review on MOBIlearn task model," *J. Interact. Media Educ.*, vol. 2015, no. 1, pp. 1–17, 2015, doi: 10.5334/jime.ap.

[21] P. Nokelainen, "An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students," *J. Educ. Technol. Soc.*, vol. 9, no. 2, pp. 178–197, 2006.

[22] P. J. Khomokhoana, B. P. J. Khomokhoana, and M. Scientiae, "Using mobile learning applications to encourage active classroom participation: Technical and pedagogical considerations," M.S. thesis, Dept. Comput. Sci. Inform., Univ. Free State, Bloemfontein, South Africa, 2011. [Online]. Available: http://scholar.ufs.ac.za:8080/xmlui/handle/11660/1082

[23] G. Zurita, N. Baloian, S. Peñafiel, and O. Jerez, "Applying pedagogical usability for designing a mobile learning application that support reading comprehension," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.*, 2019, p. 6, doi: 10.3390/proceedings2019031006.

[24] R. F. DeVellis and C. T. Thorpe, *Scale Development: Theory and Applications*. Thousand OaksSage, CA, USA: Sage Publ., 2021.

[25] C. Lacave, A. I. Molina, and M. A. Redondo, "A preliminary instrument for measuring students' subjective perceptions of difficulties in learning recursion," *IEEE Trans. Educ.*, vol. 61, no. 2, pp. 119–126, May 2018, doi: 10.1109/TE.2017.2758346.

[26] E. Knekta, C. Runyon, and S. Eddy, "One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research," *CBE Life Sci. Educ.*, vol. 18, no. 1, p. rm1, 2019.

[27] J. Cabero-Almenara, J.-J. Gutiérrez-Castillo, A. Palacios-Rodríguez, and J. Barroso-Osuna, "Development of the teacher digital competence validation of digcompedu check-in questionnaire in the university context of Andalusia (Spain)," *Sustainability*, vol. 12, no. 15, p. 6094, 2020.

[28] American Educational Research Association/American Psychological Association/National Council on Measurement in Education, *Standards for Educational and Psychological Testing*. Washington, DC, USA: Amer. Educ. Res. Assoc., 1999.

[29] M. L. Bernacki, J. A. Greene, and H. Crompton, "Mobile technology, learning, and achievement: Advances in understanding and measuring the role of mobile technology in education," *Contemp. Educ. Psychol.*, vol. 60, Jan. 2020, Art. no. 101827, doi: 10.1016/j.cedpsych.2019.101827.

[30] A. Naciri, M. A. Baba, A. Achbani, and A. Kharbach, "Mobile learning in higher education: Unavoidable alternative during COVID-19," *Aquademia*, vol. 4, no. 1, Apr. 2020, Art. no. ep20016, doi: 10.29333/aquademia/8227.

[31] S. Saikat, J. S. Dhillon, W. F. Wan Ahmad, and R. A. Jamaluddin, "A systematic review of the benefits and challenges of mobile learning during the COVID-19 pandemic," *Educ. Sci.*, vol. 11, no. 9, p. 459, Aug. 2021, doi: 10.3390/educsci11090459.

[32] S. Criollo-C, A. Guerrero-Arias, Á. Jaramillo-Alcázar, and S. Luján-Mora, "Mobile learning technologies for education: Benefits and pending issues," *Appl. Sci.*, vol. 11, no. 9, p. 4111, Apr. 2021, doi: 10.3390/app11094111.

[33] C. X. Navarro, A. I. Molina, and M. A. Redondo, "Towards a model for evaluating the usability of M-learning systems: From a mapping study to an approach," *IEEE Lat. Am. Trans.*, vol. 13, no. 2, pp. 552–559, Feb. 2015, doi: 10.1109/TLA.2015.7055578.

[34] H. Bettayeb, M. T. Alshurideh, and B. Al Kurdi, "The effectiveness of mobile learning in UAE universities: A systematic review of motivation, self-efficacy, usability and usefulness," *Int. J. Control Autom.*, vol. 13, no. 2, pp. 1558–1579, 2020.

[35] L. Punchoojit and N. Hongwarittorrn, "Usability studies on mobile user interface design patterns: A systematic literature review," *Adv. Human-Computer Interact.*, vol. 2017, pp. 1–22, Jan. 2017, doi: 10.1155/2017/6787504.

[36] N. Bevan, *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs)-Part 11: Guidance on Usability*, ISO 9241-11, Int. Org. Stand., Geneva, Switzerland, 1988.

[37] *Information Technology - Evaluation of Software Products*, ISO/IEC 14598, 1998.

[38] *Systems and Software Quality Requirements and Evaluation (SQuaRE)*, ISO/IEC 25000, 2005.

[39] *Software Engineering–Product Quality*, IEC 9126, 2001.

[40] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA, USA: Addison-Wesley, 1975.

[41] W. H. DeLone and E. R. McLean, "The delone and mclean model of information systems success: A ten-year update," *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, Apr. 2003, doi: 10.1080/07421222.2003.11045748.

[42] R. A. Alsharida, M. M. Hammood, and M. Al-Emran, "Mobile learning adoption: A systematic review of the technology acceptance model from 2017 to 2020," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 05, p. 147, Mar. 2021, doi: 10.3991/ijet.v16i05.18093.

[43] M. Al-emran, V. Mezhuyev, and A. Kamaludin, "Technology acceptance model in M-learning context: A systematic review," *Comput. Educ.*, vol. 125, pp. 389–412, Oct. 2018, doi: 10.1016/j.compedu.2018.06.008.

[44] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q.*, vol. 13, no, 3, pp. 319–340, 1989.

[45] A. Althunibat, F. Altarawneh, R. Dawood, and M. A. Almaiah, "Propose a new quality model for M-learning application in light of COVID-19," *Mob. Inf. Syst.*, vol. 2022, pp. 1–12, Mar. 2022, doi: 10.1155/2022/3174692.

[46] A. Qashou, "Influencing factors in M-learning adoption in higher education," *Educ. Inf. Technol.*, vol. 26, no. 2, pp. 1755–1785, Mar. 2021, doi: 10.1007/s10639-020-10323-z.

[47] P. Limtrairut, "Newly developed heuristics to evaluate M-learning application interface," in *Proc. Int. Conf. Inf Technol.*, 2020, pp. 197–201, doi: 10.1109/InCIT50588.2020.9310962.

[48] B. A. Kumar, M. S. Goundar, and S. S. Chand, "A framework for heuristic evaluation of mobile learning applications," *Educ. Inf. Technol.*, vol. 25, no. 4, pp. 3189–3204, 2020, doi: 10.1007/s10639-020-10112-8.

[49] B. A. Kumar and M. S. Goundar, "Usability heuristics for mobile learning applications," *Educ. Inf. Technol.*, vol. 24, no. 2, pp. 1819–1833, Mar. 2019, doi: 10.1007/s10639-019-09860-z.

[50] M. A. Almaiah and O. A. Alismaiel, "Examination of factors influencing the use of mobile learning system: An empirical study," *Educ. Inf. Technol.*, vol. 24, no. 1, pp. 885–909, Jan. 2019, doi: 10.1007/s10639-018-9810-7.

[51] N. Parsazadeh, R. Ali, M. Rezaei, and S. Z. Tehrani, "The construction and validation of a usability evaluation survey for mobile learning environments," *Stud. Educ. Eval.*, vol. 58, pp. 97–111, Sep. 2018, doi: 10.1016/j.stueduc.2018.06.002.

[52] N. Ahmad, A. Rextin, and U. E. Kulsoom, "Perspectives on usability guidelines for smartphone applications: An empirical investigation and systematic literature review," *Inf. Softw. Technol.*, vol. 94, pp. 130–149, Feb. 2018, doi: 10.1016/j.infsof.2017.10.005.

[53] C. Buabeng-Andoh, "Predicting students' intention to adopt mobile learning," *J. Res. Innov. Teach. Learn.*, vol. 11, no. 2, pp. 178–191, Nov. 2018, doi: 10.1108/JRIT-03-2017-0004.

[54] G. W. Soad, N. F. D. Filho, and E. F. Barbosa, "Quality evaluation of mobile learning applications," in *Proc. Front. Educ. Conf. (FIE)*, 2016, pp. 1–8, doi: 10.1109/FIE.2016.7757540.

[55] M. Sarrab, M. Elbasir, and S. Alnaeli, "Towards a quality model of technical aspects for mobile learning services: An empirical investigation," *Comput. Human Behav.*, vol. 55, pp. 100–112, Feb. 2016, doi: 10.1016/j.chb.2015.09.003.

[56] L. A. Clark and D. Watson, "Constructing validity: New developments in creating objective measuring instruments," *Psychol. Assess.*, vol. 31, no. 12, pp. 1412–1427, 2019.

[57] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *Mis Q.*, vol. 27, vol. 3, pp. 425–478, 2003, doi: 10.2307/30036540.

[58] C. X. Navarro, A. I. Molina, and M. Á. Redondo, "Developing a framework to evaluate usability in M-learning systems: Mapping study and proposal," in *Proc. Second Int. Conf. Technol. Ecosyst. Enhanc. Multicult.*, 2014, pp. 357–364, doi: 10.1145/2669711.2669924.

[59] *Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—System and Software Quality Models*, ISO/IEC 25010, Int. Org. Stand., Geneva, Switzerland, 2011.

[60] L. H. Sillat, K. Tammets, and M. Laanpere, "Digital competence assessment methods in higher education: A systematic literature review," *Educ. Sci.*, vol. 11, no. 8, p. 402, Aug. 2021, doi: 10.3390/educsci11080402.

[61] L. J. Simms, K. Zelazny, T. F. Williams, and L. Bernstein, "Does the number of response options matter? Psychometric perspectives using personality questionnaire data," *Psychol. Assess.*, vol. 31, no. 4, pp. 557–566, Apr. 2019, doi: 10.1037/pas0000648.

[62] D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. London, U.K.: Ravenio books, 2015.

[63] C. S. Cheah, "Factors contributing to the difficulties in teaching and learning of computer programming: A literature review," *Contemp. Educ. Technol.*, vol. 12, no. 2, May 2020, Art. no. ep272, doi: 10.30935/cedtech/8247.

[64] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[65] D. George and P. Mallery, *SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update*. Boston, MA, USA: Allyn Bacon, 2003, pp. 8–10.

[66] L. Cohen, L. Manion, and K. Morrison, *Research Methods in Education*, 8th ed., Abingdon, U.K.: Routledge, 2018, pp. 532–533.

[67] V. Zamanzadeh, M. Rassouli, A. Abbaszadeh, H. A. Majd, A. Nikanfar, and A. Ghahramanian, "Details of content validity and objectifying it in instrument development," *Nurs. Pract. Today*, vol. 1, no. 3, pp. 163–171, 2014.

[68] D. T. Campbell and D. W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychol. Bull.*, vol. 56, no. 2, pp. 81Ű-105, 1959.

[69] H. Taherdoost, "Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research," *Int. J. Acad. Res. Manag.*, vol. 5, no. 3, pp. 28–36, 2016.

[70] J. F. Hair Jr, G. T. M. Hult, C. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks, CA, USA: Sage publ., 2016.

[71] K. A. Pituch and J. P. Stevens, *Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS*. New York, NY, USA: Routledge, 2015.

[72] D. F. Fornell and C. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *J. Mark. Res.*, vol. 18, no. 1, pp. 39–50, 2016.

[73] S. Criollo-C and S. Luján-Mora, "Encouraging student motivation through gamification in engineering education," in *Proc. Mob. Technol. Appl. Internet Things*, 2019, pp. 204–211.

[74] Z. Zainuddin, S. K. W. Chu, M. Shujahat, and C. J. Perera, "The impact of gamification on learning and instruction: A systematic review of empirical evidence," *Educ. Res. Rev.*, vol. 30, Jun. 2020, Art. no. 100326, doi: 10.1016/j.edurev.2020.100326.

[75] H. Crompton, M. Bernacki, and J. A. Greene, "Psychological foundations of emerging technologies for teaching and learning in higher education," *Curr. Opin. Psychol.*, vol. 36, pp. 101–105, Dec. 2020, doi: 10.1016/j.copsyc.2020.04.011.

[76] D. Quiñones and C. Rusu, "Applying a methodology to develop user experience heuristics," *Comput. Stand. Interfaces*, vol. 66, Oct. 2019, Art. no. 103345, doi: 10.1016/j.csi.2019.04.004.

[77] J. Sauer, A. Sonderegger, and S. Schmutz, "Usability, user experience and accessibility: Towards an integrative model," *Ergonomics*, vol. 63, no. 10, pp. 1207–1220, 2020.

[78] J. Rubin and D. Chisnell, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Hoboken, NJ, USA: Wiley, 2008.

[79] R. K. Branson, G. T. Rayner, J. L. Cox, J. P. Furman, F. J. King, and W. H. Hannum, *Interservice Procedures for Instructional Systems Development: Executive Summary and Model*. Washington, DC, USA: ERIC Clearinghouse, 1975.

[80] A. Solano, C. A. Collazos, C. Rusu, and H. M. Fardoun, "Combinations of methods for collaborative evaluation of the usability of interactive software systems," *Adv. Human-Comput. Interact.*, vol. 2016, pp. 1–16, Sep. 2016, doi: 10.1155/2016/4089520.

[81] J. F. Hair, B. J. Babin, R. E. Anderson, and W. C. Black, *Multivariate Data Analysis*. Old Bridge, NJ, USA: Pearson Prentice, 2019.

Christian X. Navarro-Cota received the master's degree in computer science from the Center of Scientific Research and Superior Education, Ensenada, Mexico, in 2002, and the Ph.D. degree from the University of Castilla–La Mancha, Ciudad Real, Spain, in 2016.

He was with the Faculty of Engineering, Architecture and Design, Autonomous University of Baja California, Mexicali, Mexico, in 2001, where he is currently a Professor and a Researcher. His research interests are focused on the fields of mobile computing, human–computer interaction, ubiquitous computing, and m-learning.

Ana I. Molina received the B.E. degree in computer engineering from the University of Castilla–La Mancha (UCLM), Ciudad Real, Spain, in 2002, the Ph.D. degree in computer engineering from the UCLM in 2007, and the degree in psychology from the National University of Distance Education, Madrid, Spain, in 2019.

She is currently an Associate Professor with the UCLM. Her main lines of research focus on the application of software engineering techniques to the development of collaborative and interactive systems, the improvement of human–computer interaction processes, and the application of eye tracking techniques for the evaluation of systems, teaching materials, and notations.

Miguel A. Redondo received the Technical Engineer degree in systems of computer science from the University of Castilla–La Mancha, Ciudad Real, Spain, in 1995, the bachelor's degree in computer science from the University of Granada, Granada, Spain, in 1997, and the Ph.D. degree in computer science and engineering from the University of Castilla–La Mancha in 2002.

He is currently a Full Professor with the University of Castilla–La Mancha. His research interests include the application of software engineering techniques to the development of e-learning systems, the development and application of software engineering techniques in human–computer interaction, and the application of eye-tracking techniques in evaluation of user/learner experience.

Carmen Lacave received the M.Sc. degree in mathematics from the Complutense University of Madrid, Madrid, Spain, in 1990, and the Ph.D. degree in science from the National University of Distance Education, Madrid, in 2003.

She is currently a Professor with the University of Castilla–La Mancha, Ciudad Real, Spain. Her research is based on three lines: the application of statistical techniques and artificial intelligence (Bayesian networks) to data analysis in the fields of medicine and education; the design of experiments for the evaluation of systems and notations; and the development and evaluation of tools for teaching programming.