Journal of Modern Languages
Vol. 34, No.1 (2024)
https://doi.org/10.22452/jml.vol34no1.7

EFL Teachers' Attitudes towards Mobile Teaching Affordances: A Mokken Scale Analysis

Mona Tabatabaee-Yazdi
m.tabatabae@tabaran.ac.ir
English Department, Tabaran Institute of Higher Education, Mashhad, Iran
https://orcid.org/0000-0002-4425-3996

Aynaz Samir*
a.samir@tabaran.ac.ir
English Department, Tabaran Institute of Higher Education, Mashhad, Iran
https://orcid.org/0000-0003-3646-0159

Siavash Bakhtshirin
siavash.bakhtshirin@nisantasi.edu.tr
Foreign Languages Department, Lecturer, Nişantaşi University Turkiye
https://orcid.org/0009-0004-5984-3026

Seyedeh Maryam Tabatabaeiyazdi
s.tabatabaeiyazdi@bbnairlines.aero
Executive Assistant to Managing Director, In charge of the English Language Assessment of the employees, BBN Airlines Turkiye
https://orcid.org/0000-0001-5041-113X

Abstract

In contemporary education settings, the expeditious advancement of mobile devices, wireless communications and network infrastructures has emerged as a crucial issue. Although mobile technology capabilities and affordances have attracted considerable attention, there is a lack of a valid scale for evaluating mobile teaching affordances. Therefore, this research sought to develop and validate a 36-item inventory of mobile teaching affordances. To accomplish this objective, 204 EFL teachers were selected for this research based on a convenience sampling method. The data was analysed using the Mokken scale analysis. The study results indicated that 10 out of 36 items should be deleted to optimise the utilisation of the data. Consequently, the study created a valid and reliable

*Corresponding author

26-item inventory that could be used to measure mobile teaching affordance in EFL or ESL teaching contexts. The research findings suggested that the newly created scale is an invaluable instrument for evaluating the perspectives of EFL teachers regarding the potential benefits of mobile teaching. Furthermore, it can effectively pinpoint areas where teachers may require further development in their understanding of and proficiency in utilising mobile teaching methods. These revelations can also act as a basis for developing instructional programmes aimed at empowering EFL teachers to seamlessly incorporate mobile devices into their pedagogical approaches, thereby enhancing their overall effectiveness. The study's findings can be used by teachers, learners and researchers in educational systems.

Keywords: Affordances, EFL Teachers' Attitudes, IRT Models, Mobile Assisted Language Learning, Mokken Scale Analysis

1. Introduction

The expeditious progress of portable gadgets, mobile devices, wireless communications and network technologies has become integral to the university environment for both teachers and students. Accordingly, this evolution, driven by the internet and the Web, has significantly impacted various aspects of human life. Many scholars and academics have posited that online learning surpasses face-to-face instruction in the realm of education. One of the significant consequences of this rapid evolution is the seamless connection of teachers and students to other devices at any time and place, facilitated by the internet and wireless technologies (Ibrahim & Kadiri, 2018). This connectivity has engendered a greater flexibility in the realm of education as mobile devices particularly smartphones and tablets facilitate the recording, transfer and retrieval of information from any given location (Welsh et al., 2015). The accessibility of a diverse range of mobile devices combined with the portability and versatile nature of these devices has strengthened their capacity to enhance student achievements. Additionally, the utilisation of mobile devices nurtured a more student-centred approach to education – empowering students to take greater responsibility for learning, processing, and applying data. Moreover, mobile technology has not solely provided assistance to the process of student acquisition of knowledge, but has additionally assisted in fostering enhanced involvement between educators and learners, thereby personalising the educational experience (MacCallum et al., 2017).

In light of these developments, this study endeavoured to construct and validate an inventory of EFL teachers' attitudes towards *Mobile Teaching Affordances* by employing the Mokken scale analysis to investigate mobile device affordances that are used and considered effective by Iranian EFL teachers in their professional lives. Consequently, the research aimed to tackle the subsequent research inquiries:

RQ1. Does the developed inventory of mobile teaching affordances in the EFL context exhibit construct validity?

RQ2. Is the developed inventory of mobile teaching affordances in the EFL context reliable?

2. Literature Review

A review of the existing literature reveals limited investigation of mobile teaching affordances. Rostami (2021) has examined mobile learning affordances from the learners' perspective. Therefore, this study seeks to address this gap by designing and validating an inventory of Iranian EFL teachers' attitudes towards mobile teaching affordances. To achieve this, several empirical studies on mobile language education were examined.

2.1 Teaching Affordances

The term 'affordance' as defined in the Oxford English Dictionary (n.d.), refers to the qualities, properties, or resources available in the environment to an object. Gibson (1979/2014, as cited in Egessa et al., 2021) developed this concept linking it to both the environment and human beings and suggesting that the environment influences behaviour. It is also indicated that interaction between the environment and an actor represents an affordance, which is not dependent on the ability of the actor or environment properties.

In the context of education, Kirschner et al. (2004) provided an exposition on educational affordances, defining them as the inherent characteristics of an artefact that ascertains the possibility and manner in which a distinct learning behaviour could manifest within a specific setting. These affordances represent the connections between the attributes of an educational approach and the traits of a student that facilitate specific forms of learning (Kirschner et al., 2004). For instance, Bahari and Salimi (2021) asserted that the affordances offered by mobile technology have the capacity to enhance the motivation of L2 learners and reduce their anxiety. Similarly, Gunter and Braga (2018) highlighted the original affordances and applications of tablets and smartphones such as speakers, cameras and microphones, which are evident to the users and can be easily utilised. When users are familiar with the applications of mobile technology, it leads to the emergence of affordances, demonstrating the potential possibilities within the environment (Gunter & Braga, 2018). This study focuses on four important affordances of mobile devices that can be beneficial in promoting English teaching among Iranian EFL teachers.

2.2 Connectivity and mobile technology

Reeves et al. (2017) conducted an analysis of the connectivity of mobile technology, defining it as the capacity to swiftly acquire information immediately through the internet. This enables people to stay updated by accessing online resources such as library, course content and emails**.** Mobile technology allows for the gathering of and response to data as well as the storage of essential data through capabilities such as taking photos, and recording audio and video files (Parsons et al., 2016). The small size and portability of mobile devices enable their use in various contexts like being able to support ubiquitous learning and to facilitate the transfer of knowledge between different environments (Naismith et al., 2004; Sharples, 2002). Ibrahim and Kadiri (2018) highlighted that mobile phones enable teachers to assess and monitor students' learning, correct mistakes via text messages and promote critical thinking.

2.3 Mobile Teaching Affordances

The utilisation of Mobile Technologies (MT) for educational purposes is commonly known as M-learning, allowing for education without the constraints of a fixed location or schedule (MacCallum & Kinshuk, 2006). Nyiri (2002) describes M-learning as learning that occurs during person-to-person mobile communication. It entails the use of, Personal Digital Assistants (PDAs), iPods and phones. The realm of mobile learning and teaching encompass a wide range of activities such as web and podcasting to applications that convert audio into text or text into audio (Gaudry-Perkins & Dawes (2011). Additionally, it includes the utilisation of simple Short Message System (SMS) messaging, live classroom sessions, Multimedia Messaging Service (MMS), and the provision of enhanced learning experiences through educational videos, exercises in logical reasoning and problem-solving aptitude games. The advantage of mobile technologies is mobility, which applies to both the physical structure of the device itself and the participants involved.

Thus, mobile teaching, also known as M-learning, encompasses various concepts such as mobile learning, ubiquitous learning, learning anytime, anywhere, and learning manually, with each one holding different meanings for different people or communities. According to Kukulska-Hulme (2012), Mobile Assisted Language Learning (MALL) refers to the utilisation of smartphones and other portable technologies in contexts that provide advantages through portability and situated learning.

Recently, Gutiérrez-Colón et al. (2023) conducted a comprehensive review of implementation studies focusing on enhancing second language reading comprehension by means of mobile devices. The study provided a historical background of Mobile-Assisted Language Learning (MALL) applications and reviewed selected studies between 2012 and 2017 by offering a synopsis of conference proceedings, book chapters, and primary journal articles that examine the utilisation of tablets and mobile devices to enhance second language reading comprehension. Additionally, Li (2023) highlighted the benefits of MALL for EFL learners' listening skills. The findings suggested that MALL outperformed traditional methods when it comes to teaching listening.

Accordingly, Togaibayeva et al. (2022) investigated the influence of different characteristics of mobile learning on the satisfaction of students, the perceived utility of employing mobile technologies in foreign language acquisition, and students' educational success. The investigation unveiled those individual elements of mobile learning including the impetus to engage in portable learning, the pertinence of information disseminated through mobile learning to the needs of students, the omnipresence of mobile learning, and self- assurance, among others, significantly affected students' perception of foreign language learning, their perception of the usefulness of portable learning, and their scholastic achievements.

The study by Chen and Tsai (2021), which researched teachers' perceptions of mobile technology, involved individual interviews with 25 Taiwanese educators enlisted from three primary schools. These institutions actively integrated Information and Communications Technology (ICT) into their curricula, granting their students access to mobile devices, wireless communication technology, and interactive learning environments to bolster teaching and learning. Additionally, each student was equipped with a mobile device throughout their time at school. The interview data included seven qualitative classifications (7-T) of mobile learning. The results underlined the advantages that mobile technology brought to student-centred learning, equipping teachers with a more profound understanding of sophisticated technology-enhanced learning and thereby, facilitating them to design student-centred learning activities.

Furthermore, the scholarly inquiry conducted by Hoi and Mu (2020) delved into the acceptability of mobile devices among Vietnamese students pursuing higher education. The results showed that students preferred teachers' instruction regarding which mobile resources to use in language learning in both in-class and out-of-class contexts more than actual demonstration of mobile-assisted language learning activities that took place during an in-class session.

In a separate study, García-Martínez et al. (2019) conducted a review of the favourable impact of mobile devices on the learning of university students. By surveying six articles, the authors summarised that Mobile learning (M-learning) significantly enhanced students' learning.

Alaghbary (2021), who carried out a study in Saudi Arabia to examine and compare teachers' and learners' utilisation of mobile phones to facilitate learning and teaching, found a positive correlation amidst the possibilities provided by mobile technology and its capacity to elevate the art of instruction and acquisition. However, there is a negative correlation between the actual utilisation of mobile-assisted technology by learners and teachers to bolster pedagogical practices, acquaintance with MALL tools, software and activities, and perceptions about the limitations of MALL standardisation in the language classroom.

In the same vein, Ibrahim and Kadiri (2018) delved into the untapped possibilities of harnessing the power of mobile technology as a valuable educational tool for the enhancement of English language acquisition. The study concluded that M-learning is learner-centred, collaborative and cooperative. It increased participative learning through non-traditional teaching methods such as web networks and forums where students were allowed to post their opinions and share information. M-learning provided learners with the chance to use real English language which was the basis for building one's learning. The inclusion of M-learning in English instruction could prepare learners for the benefits of the digital age.

2.4 The Mokken Scale Analysis

The Mokken Scale Analysis (MSA) which represents a non-parametric approach within the realm of item response theory (IRT) is designed to evaluate unidimensional scales that are dichotomous (binary) or polytomous (ordinal) in nature (Mokken, 1971). When data fits the nonparametric item response theory (NIRT) models, it suggests that the testers can be ranked based on their aggregate scores (Baghaei, 2021; Tabatabaee-Yazdi et al., 2021). Thus, a greater score on an item signifies a higher degree of a characteristic under assessment. Several researchers (Firoozi, 2021; Meijer & Baneke, 2004; Meijer et al., 2015; Wind, 2017) contended that in light of basic presumptions of item response theory (IRT), raw scores represent ordinal scale data, signifying that the ranking of the examinees can be ascertained, rather than their disparities in their strengths or shortcomings. Nevertheless, for unprocessed scores to be considered ordinal, the patterns of responses must adhere to a principle called transitivity, which implies that if a respondent correctly responds to a more demanding question, they should also have correctly answered a less challenging item (Sijtsma & Molenaar, 2002, as cited in Baghaei, 2021).

MSA is utilised in both the confirmatory and exploratory approaches. In the confirmatory approach, it is postulated that multiple items come together to constitute a scale, whereas in the exploratory approach, a subset of items is examined to determine whether they form one or possibly two or more scales. Both approaches rely on identical standards based on two theoretical frameworks the Double Monotonicity Model (DMM) and the Monotone Homogeneity Model (MHM) (Mokken, 1971). DMM implies an invariant item ordering while the MHM does not (Sijtsma & Van Der Ark, 2017).

3. Method

3.1 Participants

In previous studies, the sample size for MSA varied between 133 and 15,022. Some studies used small sample sizes for MSA and yielded satisfactory results (Adler & Brodin, 2011; Lee et al., 2017; Tabares et al., 2021). For instance, Adler and Brodin (2011) employed a sample size of 133 to examine the psychometric characteristics of the Affective Self-Rating Scale (AS-18). Lee et al. (2017) used a sample size of 214 to assess the Chinese version of the Hamilton Depression Rating Scale. Using a sample size of 214, Chou et al. (2017) also used MSA to examine the unidimensionality of the Depression and Somatic Symptoms Scale (DSSS). In another study, Tabares et al. (2021) used a sample of 193 participants to explore the psychometric qualities of the Columbia Suicide Severity Rating Scale (CSSRS) in terms of between- and within-person measurement dimensions. Based on a simulation study, Straat et al. (2014) suggested that the optimal number of respondents for a successful MSA ranges between 250 and 500 when the quality of the items is exceptionally high. Moreover, Watson et al. (2018) showed that a broad range of sample sizes, surpassing the threshold of 250 to over 1,000, boasted an ample magnitude to conduct an MSA experiment. Furthermore, according to Cohen (1988) and Soper (2020), it was determined that the absolute minimum total sample size and the minimum sample size per group required for a comprehensive two-tailed t-test investigation would be 128 and 64 respectively, with an esteemed probability level of 0.05, an effect size (Cohen's *d*) of 0.5, and an unwavering power level of 0.8. Besides, Wright and Stone (1979) stated that the closer the dataset is to the centre of the distribution of the 100, the more believable the results.

Therefore, a comprehensive group of 204 EFL teachers were selected through convenience sampling. They were from both genders (138 [67.7%] female, 66 [32.4%] males), different age groups (*M*= 29.91, *SD*= 7.78), and different fields of study (English Teaching [56%], English Translation [21%], English Literature [15%], and Applied Linguistics [8%]). The sample size of 204 EFL teachers in this study was determined based on the requirements for MSA studies and the results of previous research. While some studies have used small sample sizes for MSA and yielded satisfactory results, this study chose a larger sample to ensure that the sample accurately represents the population and encompasses the entire spectrum of disparities in the measurement process. This sample size is larger than the minimum required for a two-tailed t-test study as suggested by Cohen (1988) and Soper (2020), and it exceeds the minimum sample size recommendations for MSA studies provided by Straat et al. (2014) and Watson et al. (2018). Convenience sampling was chosen for this study due to its feasibility and cost-effectiveness. It allowed us to select participants based on their availability and willingness to take part, which was important given the specific expertise required for the study. To mitigate potential biases during participant selection, we ensured that the sample encompassed a diverse range of genders, ages and fields of study. This approach significantly increased the likelihood of a representative sample that fully encapsulates the entire spectrum of disparities in the measurement process.

3.2 Instruments

3.2.1 Teachers' Attitudes towards Mobile Teaching Affordances Inventory

To achieve the objectives of this study, 59 items, categorised into four constructs, pertaining to mobile teaching affordances used by teachers, were selected from the literature. Subsequently, 10 specialists in the realm of English teaching evaluated the content validity of the inventory. Following the feedback from experts, some modifications and changes were applied (i.e., 18 items were removed, 11 items were merged into 4 items, and 2 items were added). After the necessary modifications, a 36-item inventory was designed, incorporating 4 dimensions of assessments (items 33-36), connectivity (items 1-16), mobility and interaction with others (items 26-32), and context-sensitivity (items 17 to 25), and utilising a five-point Likert scale (see Appendix A).

3.2.2. Scale Development of the Inventory

The inventory items were selected by reviewing related literature and seeking experts' opinions in the realm of English language teaching. Gunter and Braga (2018) proposed ten items (items 1, 2, 3, 4, 5, 7, 12, 13, 21, 28) related to mobile teaching affordances, including accessing online libraries, distributing course materials to students, checking cultural events, checking the news of the institution and their portal, sending homework assignments to students, reading electronic books, searching information, filling inventory, and arranging teaching appointments.

Further, Keskin and Metcalf (2011) designed seven items (items 6, 9, 22, 27, 29, 30 and 36) related to mobile teaching affordances, including checking emails, sending videos or pictures to help the initiation of a discussion, making contact with colleagues, asking students questions by sending messages, getting answers to questions, and checking students' homework assignments.

Parsons et al. (2016) designed eight items (items 8, 14, 17, 18, 19, 24, 31, and 32) related to mobile teaching affordances, including sending voice or video files to students, collecting data using different ways such as Google Docs/Forms, recording classes for students to review in the future, taking photos of class presentations, saving different types of course content to use in the future, attending online classes on different platforms through mobile technology, enhancing communication in the classroom, and enhancing collaborative learning in the classroom. Gaudry-Perkins and Dawes (2011) proposed one item (item 10) related to mobile teaching affordances, which is sharing beliefs, ideas, experiences, and viewpoints.

Items 15, 25 and 33 were designed according to Ibrahim and Kadiri (2018) who believed that teachers can usually use mobile phones to download videos and play them to students, utilise mobile dictionaries to find out word meanings, and assess and monitor students' learning. MacCallum et al. (2017) proposed five items (items 16, 20, 23, 26, and 34) related to mobile teaching affordances, which are sending posts/movies to Instagram, YouTube, or Facebook, taking notes in classes, accessing Learning Management System (LMS), learning/teaching with no restriction to a certain time or a specific place, and taking tests and quizzes using Google Docs. Finally, Sharples (2002) proposed an item (item 35) related to mobile teaching affordances, which is providing students with feedback and error corrections.

3.3 Procedure

The present study employed a quantitative-descriptive research design to develop and validate teachers' attitudes towards mobile teaching affordances via an inventory. Fifty-nine items, categorised into four constructs, pertaining to mobile teaching affordances utilised by teachers were drawn from the literature. A group of 10 specialists in the realm of English teaching evaluated the content validity of the inventory. Following feedback from experts, several modifications were made, including the removal of 18 items, the consolidation of 11 items into 4, and the addition of 2 items. Next, the revised 36-item inventory was distributed electronically using a Google Form link to collect data due to the COVID-19 pandemic. Data were collected from August 2021 to December 2021 using a convenience sampling method. It took approximately 10 minutes to complete the inventory. The data underwent analysis using SPSS version 24.0 (IBM Corp., 2016) and R Software version 4.1.0 (R Core Team, 2021), and the reliability and construct validity of the inventory were assessed through the Mokken Scale Analysis.

The researchers of this study bore a significant responsibility for ascertaining the accuracy, authenticity, and reliability of the collected data. Participants were ensured that their responses were anonymised and only accessible to those directly involved. Participation was voluntary, and participants were provided with sufficient information about the research objectives to make an informed decision regarding their involvement.

4. Results

4.1 Data Presentation using the Mokken Scale Analysis

The following sections present the data based on the Mokken Scale Analysis.

4.1.1 Scalability Coefficients

Data analysis was conducted using the Mokken package version 3.0.6 (van der Ark et al., 2021).. To determine whether the items of the scale could come together to create a Mokken scale, we conducted a meticulous examination of the scalability coefficients for all items in the scale (H), as well as the scalability coefficient for each individual item in the scale (Hi), and the scalability coefficients for each item-pair (Hij). According to the proposed standards by Mokken (1971), a scale is deemed feeble if the scalability coefficient (H) falls within the range of 0.30 to 0.40, moderate if it falls within the range of 0.40 to 0.50, and strong if it is greater than or equal to 0.50. Additionally, the magnitudes of Hij transcend the realms of zero or the non-negative spectrum. Should the coefficient Hj fail to surpass the threshold of 0.30, a meticulous evaluation or elimination of the items is warranted. Nevertheless, in the event that the coefficient Hj meets or exceeds the minimum threshold of 0.30, the selection and amalgamation of the items will ultimately culminate in the creation of a Mokken scale. The outcomes of the inter-item scalability coefficients (Hij) demonstrated a sublime portrayal of all item pairs, with a delightful range extending from 0.31 to 0.78. These findings imply a harmonious and affirmative correlation between an item and the elusive attribute being assessed. The results of scalability coefficients for the items as well as whole scale, including their established discrepancies, are presented in Table 1. The scalability coefficients (Hi) for the items ranged from 0.26 to 0.43. With the exception of items 1, 2 and 3, the scalability coefficients for the other items exceed 0.30. The scalability coefficient for the whole scale was 0.36 (SE = 0.041), indicating a weak scale.

Table 1. Item Scalability Coefficients of the Scale Items, the Whole Scale, and their Standard Errors

Items	Scalability Coefficients	Standard Errors (SE)
1	0.246	0.061
2	0.261	0.069
3	0.264	0.062
4	0.304	0.062
5	0.307	0.061
6	0.340	0.055
7	0.384	0.053
8	0.420	0.055
9	0.381	0.053
10	0.358	0.046
11	0.353	0.047
12	0.360	0.046
13	0.398	0.062
14	0.414	0.049
15	0.382	0.052
16	0.324	0.051
17	0.398	0.040
18	0.316	0.050
19	0.362	0.055
20	0.382	0.049
21	0.350	0.057
22	0.370	0.044
23	0.358	0.053
24	0.394	0.045
25	0.375	0.067
26	0.415	0.047
27	0.363	0.059
28	0.406	0.055
29	0.413	0.046
30	0.360	0.057
31	0.396	0.048
32	0.401	0.048
33	0.409	0.044
34	0.403	0.044
35	0.422	0.048
36	0.432	0.046
Scale	0.368	0.041

4.1.2 Automated Item Selection Procedure (AISP)

To assess the unidimensionality of the scale, MSA was employed, which offers an automated item selection procedure (AISP). Unidimensionality implies that each item within the scale should gauge a solitary underlying characteristic. The ASIP analysis aids in the selection of an assortment of (scalable) items that gauge the identical underlying characteristic, and it adheres to the monotone homogeneity model, which is similar to exploratory factor analysis (EFA). AISP, a method for splitting the data into several subscales that meet the MSA criteria, can also identify and remove non- or low-discriminating items (Baghaei, 2021; Sijtsma & van der Ark, 2017). The outcomes of AISP for the scale are presented in Table 2. As per Sijtsma and van der Ark (2017), the uppermost value of 0.30 signifies the lower threshold for the scalability coefficient during the creation of scales, as mentioned by the value of 0.30 at the top, where '0' denotes unscalable items, '1' signifies items belonging to scale 1, and '2' represents items belonging to scale 2. The AISP revealed two scales and one unscalable item out of the 36 items, with items 1 and 2 forming a short scale, and the remaining items forming a unidimensional scale.

Table 2. The Outcomes of Automated Item Selection Procedure (AISP) for the Scale

Items	**Dimensions from AISP c = 0.30**	**Items**	**Dimensions from AISP c = 0.30**
1	2	19	1
2	2	20	1
3	0	21	1
4	1	22	1
5	1	23	1
6	1	24	1
7	1	25	1
8	1	26	1
9	1	27	1
10	1	28	1
11	1	29	1
12	1	30	1
13	1	31	1
14	1	32	1
15	1	33	1
16	1	34	1
17	1	35	1
18	1	36	1

Note: c = 0.30 is the cut-off value or lower bound of the scalability coefficient

4.1.3 Monotonicity

Monotonicity is a crucial assumption in MSA, which states that as the level of individuals' ability (e.g., θ) increases, the likelihood of providing an accurate response or endorsing a response alternative should increase. Table 3 presents the analysis of the monotonicity assumption, where the second column (#ac) indicates the overall count of operational sets of relaxation score groups used to test manifest monotonicity. The third column (#vi) illustrates the total number of infractions, and the fourth column (#vi/#ac) presents the mean count of infractions per operational set. Columns five (maxvi) and six (sum) represent the utmost infringement and the accumulation of all transgressions respectively. The seventh column (sum/#ac) demonstrates the mean violation per functioning pair, while columns eight (zmax) and nine (#zsig) separately exhibit the utmost test statistic and the count of notable infractions. The last column (crit) shows a weighted sum of elements such as 'item H', '#ac', and so on. A high value of 'crit' shows poor items. According to Molenaar and Sijtsma (2000), items with 'crit' values less than 0.40 demonstrate adherence to the monotonicity hypothesis, while values exceeding 0.40 indicate a breach of monotonicity. As shown in Table 3, all items of the scale meet the assumption of monotonicity, supporting the arrangement of testers in terms of their total scores. Figure 1 illustrates the visual analysis of monotonicity assumptions for item 2 of the scale, where each plot includes two parts. The left part shows the Item Step Response Function (ISRF), which represents the likelihood of endorsing a certain category across the latent trait θ, whereas the right part showcases the Item Response Function (IRF) for the overall item, which effectively delineates the correlation between the concealed characteristic and the items or response alternatives. The plots confirm the numerical values of monotonicity analysis and illustrate that ISRFs and IRFs are non-decreasing.

Table 3. The Results of the Monotonicity Assessment

Items	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	Zmax	#zsig	Crit
1	4	0	0	0	0	0	0	0	0
2	3	0	0	0	0	0	0	0	0
3	4	0	0	0	0	0	0	0	0
4	4	0	0	0	0	0	0	0	0
5	4	0	0	0	0	0	0	0	0
6	2	0	0	0	0	0	0	0	0
7	4	0	0	0	0	0	0	0	0
8	3	0	0	0	0	0	0	0	0
9	3	0	0	0	0	0	0	0	0
10	3	0	0	0	0	0	0	0	0
11	3	0	0	0	0	0	0	0	0
12	4	0	0	0	0	0	0	0	0
13	2	0	0	0	0	0	0	0	0
14	3	0	0	0	0	0	0	0	0
15	4	0	0	0	0	0	0	0	0
16	4	0	0	0	0	0	0	0	0
17	4	0	0	0	0	0	0	0	0
18	4	0	0	0	0	0	0	0	0
19	4	0	0	0	0	0	0	0	0
20	4	0	0	0	0	0	0	0	0
21	4	0	0	0	0	0	0	0	0
22	4	0	0	0	0	0	0	0	0
23	4	0	0	0	0	0	0	0	0
24	4	0	0	0	0	0	0	0	0
25	1	0	0	0	0	0	0	0	0
26	3	0	0	0	0	0	0	0	0
27	2	0	0	0	0	0	0	0	0
28	3	0	0	0	0	0	0	0	0
29	3	0	0	0	0	0	0	0	0
30	4	0	0	0	0	0	0	0	0
31	4	0	0	0	0	0	0	0	0
32	4	0	0	0	0	0	0	0	0
33	4	0	0	0	0	0	0	0	0
34	4	0	0	0	0	0	0	0	0
35	3	0	0	0	0	0	0	0	0
36	3	0	0	0	0	0	0	0	0

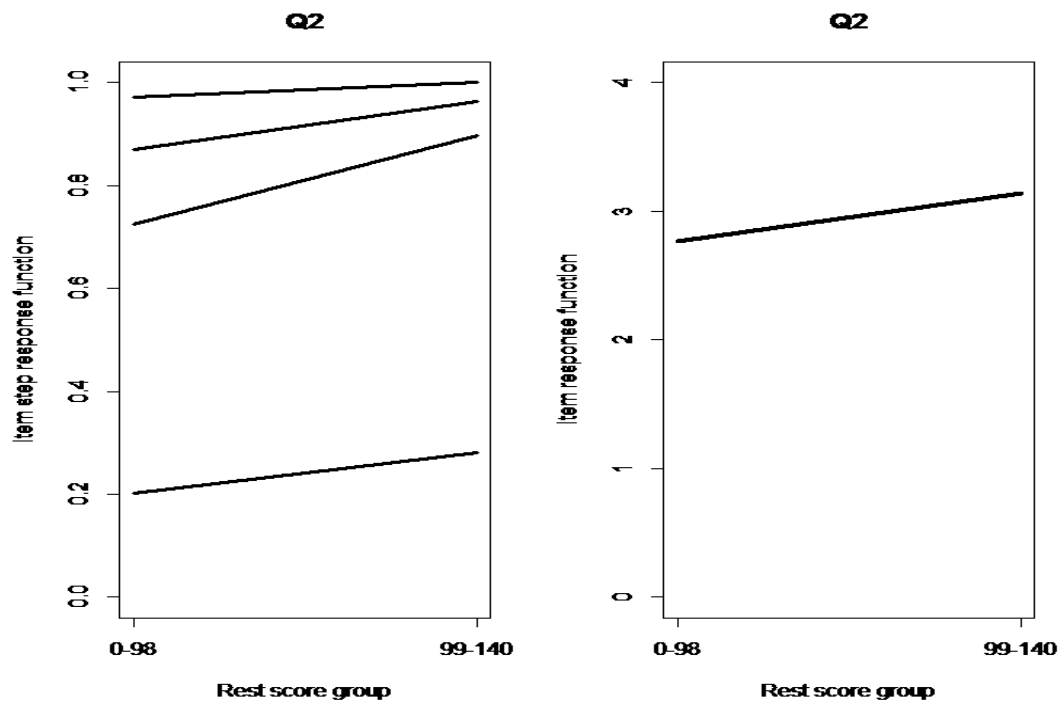

Figure 1. Monotonicity Plots for Six Items of the Scale

4.3.4 Invariant Item Ordering (IIO)

The use of invariant item ordering (IIO) or non-intersection of IRFs serves to ascertain whether the ordering of items remains consistent for all respondents, regardless of their trait levels. IIO pertains to items that possess identical levels of difficulty in terms of their ordering, regardless of the trait level. This aids practitioners in arranging items based on their difficulty or prevalence, and this significantly eases the understanding of examination results in the utilisation of stock and assessments (Sijtsma & van der Ark, 2017). More importantly, IIO imparts a heightened significance to comparisons between individuals based on their overall scores. That is to say, if a respondent possesses a superior total score in comparison to another individual, they will exhibit a greater likelihood of having triumphed in an item or having endorsed higher categories. Likewise, two individuals with identical total scores are more prone to having responded correctly to the same items or to having endorsed the same categories. Overall IIO can be conducted using the HT coefficient (Sijtsma & van der Ark, 2017). The employed methodology for interpreting HT (Ligtvoet et al., 2010) is based on a rule of thumb. When HT is < 0.3, it signifies insufficient or inaccurate IIO. In the range of $0.3 \le HT < 0.4$, it suggests a weak IIO. Moving on, when the value is $0.4 \le HT < 0.5$, it implies a moderate IIO. Lastly, when HT is ≥ 0.5, it indicates a strong IIO. The IRFs of two items can also be compared graphically to determine if they intersect significantly. The analysis of the scale's IIO is summarised in Table 4. The second column (ItemH) represents the scalability coefficient for each item. The third column (#ac) indicates the indicates the total number of active pairs, and columns four (#vi) and five (#vi/#ac) shows respectively demonstrate the total number of violations and the average number of violations per active pair. The sixth column (maxvi) discloses the maximum violation. Columns seven (sum) and eight (sum/#ac) display the sum of all violations and the average violation per active pair, respectively. Columns nine (tmax) and ten (#tsig) present the maximum test statistic and and the number of significant violations. The last column (crit) is a weighted sum of the other elements such as ‘item H’, and ‘#ac’. The ‘crit’ value can be used to indicate the effect size of the IIO violation, with a high value indicating poor items. As suggested by Sijtsma and Molenaar (2002) and van Schuur (2011), the pinnacle of perfection is embodied in a value of 0. When Crit falls below 40, a mere whisper of transgression is detected. A range of 40 to 80 on the Crit scale hints at a violation that may be deemed trivial, yet the items in question must undergo scrutiny. Lastly, a Crit score of 80 or higher is indicative of an egregious or profound violation. In Table 4, the HT value is 0.154, suggesting inaccurate item ordering. Twenty-four (24) items have violated IIO (column #vi), but only ten items (e.g., 1, 2, 3, 4, 5, 12, 16, 26, 29, 36) have significant violations (column #tsig). For example, item 2 has 11 violations, meaning the IRF for this item intersects with the IRF of eleven other items, but only five of these violations are significant. Similarly, the IRF for item 3 intersects with the IRF of six other items, but only one of them is significant. The backward selection method was used to remove items which violate IIO. If there are an equal number of violations for two or more items, then the item with the lowest scalability is removed (Ligtvoet et al., 2010). As demonstrated in Table 8, the results of backward selection showed that ten items (e.g., 1, 2, 3, 4, 5, 12, 16, 26, 29, and 36) have violated IIO and should be removed. To delete items, one item at a time was removed in iterative steps because “IIO violations of other items may be influenced by the inclusion or exclusion of any particular item (Stochl et al., 2012).

Table 4. The Summary of IIO Analysis for the Scale

Items	ItemH	#ac	#vi	#vi/#ac	Maxvi	sum	sum/#ac	tmax	#tsig	Crit
25	0.38	37	0	0.00	0.00	0.00	0.0000	0.00	0	0
13	0.40	37	0	0.00	0.00	0.00	0.0000	0.00	0	0
8	0.42	35	0	0.00	0.00	0.00	0.0000	0.00	0	0
27	0.36	36	0	0.00	0.00	0.00	0.0000	0.00	0	0
6	0.34	37	0	0.00	0.00	0.00	0.0000	0.00	0	0
21	0.35	36	0	0.00	0.00	0.00	0.0000	0.00	0	0
7	0.38	38	4	0.11	0.25	0.69	0.0183	1.47	0	66
30	0.36	35	1	0.03	0.13	0.13	0.0037	0.78	0	24
14	0.41	35	3	0.09	0.27	0.68	0.0193	1.56	0	67
4	0.30	38	4	0.11	0.28	0.97	0.0256	2.45	1	100
28	0.41	41	1	0.02	0.18	0.18	0.0043	1.07	0	29
9	0.38	36	2	0.06	0.17	0.33	0.0092	1.04	0	40
5	0.31	37	4	0.11	0.20	0.67	0.0180	1.81	1	80
36	0.43	37	5	0.14	0.36	1.19	0.0322	2.45	3	125
15	0.38	35	2	0.06	0.28	0.44	0.0125	1.49	0	57
29	0.41	37	4	0.11	0.38	0.91	0.0246	2.16	1	103
19	0.36	36	2	0.06	0.18	0.32	0.0090	1.08	0	41
26	0.42	38	2	0.05	0.35	0.56	0.0147	2.12	1	81
2	0.26	36	11	0.31	0.38	2.61	0.0725	2.16	5	209
16	0.32	36	3	0.08	0.28	0.64	0.0178	1.86	1	84
3	0.26	38	6	0.16	0.28	1.30	0.0341	1.71	1	114
12	0.36	36	5	0.14	0.37	1.01	0.0280	1.81	3	122
1	0.25	35	5	0.14	0.37	0.98	0.0279	1.81	1	115
11	0.35	36	1	0.03	0.13	0.13	0.0035	1.32	0	27
35	0.42	35	1	0.03	0.12	0.12	0.0035	0.69	0	20
23	0.36	35	0	0.00	0.00	0.00	0.0000	0.00	0	0
31	0.40	36	0	0.00	0.00	0.00	0.0000	0.00	0	0
24	0.39	36	1	0.03	0.25	0.25	0.0068	1.23	0	41
22	0.37	37	1	0.03	0.12	0.12	0.0033	0.68	0	22
18	0.32	36	2	0.06	0.25	0.37	0.0103	1.23	0	53
20	0.38	35	0	0.00	0.00	0.00	0.0000	0.00	0	0
34	0.40	36	1	0.03	0.13	0.13	0.0035	0.76	0	22
32	0.40	37	1	0.03	0.13	0.13	0.0034	0.76	0	22
33	0.41	38	0	0.00	0.00	0.00	0.0000	0.00	0	0
17	0.40	36	0	0.00	0.00	0.00	0.0000	0.00	0	0
10	0.36	35	0	0.00	0.00	0.00	0.0000	0.00	0	0

*HT = 0.154

Table 5. The Results of the Backward Item Selection Method

Items	Step1	Step2	Step3	Step4
25	0	0	0	0
13	0	0	0	0
8	0	0	0	0
27	0	0	0	0
6	0	0	0	0
21	0	0	0	0
7	0	0	0	0
30	0	0	0	0
14	0	0	0	0
4	1	1	1	0
28	0	0	0	0
9	0	0	0	0
5	1	1	1	0
36	3	2	2	NA
15	0	0	0	0
29	1	0	0	0
19	0	0	0	0
26	1	0	0	0
2	5	NA	NA	NA
16	1	0	0	0
3	1	1	0	0
12	3	2	NA	NA
1	1	1	0	0
11	0	0	0	0
35	0	0	0	0
23	0	0	0	0
31	0	0	0	0
24	0	0	0	0
22	0	0	0	0
18	0	0	0	0
20	0	0	0	0
34	0	0	0	0
32	0	0	0	0
33	0	0	0	0
17	0	0	0	0
10	0	0	0	0

The quality of the 26-item scale was analysed after removing the items violating the IIO. Table 6 summarises the analysis of the 26-item scale. The total scalability of the 26-item scale is 0.395, and all item pairs are positive. The item coefficients of scalability for the item surpass the threshold of 0.30. The results of the IIO and backward selection method also show that although five items have violated IIO, they are not significant. This indicates that after removing the ten items, the Double Monotonicity Model (DMM) and the Monotone Homogeneity Model (MHM) fit harmoniously with the data even after eliminating ten items.

Table 6. The Summary of the Analysis for the 26-item Scale

Items	Dimensions from AISP	Items	#ac	#vi	#vi/#ac	Maxvi	sum	sum/#ac	tmax	#tsig	Crit	Step1
25	1	0.40	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
13	1	0.40	28	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
8	1	0.44	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
27	1	0.36	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
6	1	0.34	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
21	1	0.37	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
7	1	0.40	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
30	1	0.38	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
14	1	0.42	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
28	1	0.41	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
9	1	0.41	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
15	1	0.41	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
19	1	0.39	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
11	1	0.36	25	1	0.04	0.12	0.12	0.0050	0.75	0	26	0
35	1	0.44	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
23	1	0.37	25	2	0.08	0.13	0.26	0.0104	1.22	0	41	0
31	1	0.41	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
24	1	0.41	25	3	0.12	0.21	0.53	0.0214	1.22	0	63	0
22	1	0.38	25	1	0.04	0.21	0.21	0.0083	1.06	0	40	0
18	1	0.34	26	1	0.04	0.19	0.19	0.0074	1.00	0	38	0
20	1	0.40	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
34	1	0.41	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
32	1	0.41	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
33	1	0.42	26	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
17	1	0.42	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0
10	1	0.37	25	0	0.00	0.00	0.00	0.0000	0.00	0	0	0

*H = 0.395 (se = 0.040)

4.3.5 Reliability

The reliability of the finalised scale was scrutinised using four distinct reliability coefficients: Lambad-2 (Guttman, 1945), Mokken scale (MS) reliability or ρ (Mokken, 1971), Cronbach's alpha (Cronbach, 1951), and the latent class reliability coefficient (LCRC) (van der Ark et al., 2011). As demonstrated in Table 7, all the values exceeded 0.90, indicating the high reliability of the scale.

Table 7. Reliability Indices for the 26-item Scale

Reliability Index	MS	Alpha	Lambad-2	LCRC
Value	0.941	0.939	0.941	0.941

5. Discussion

The primary objective of the current investigation was to construct and validate a scale for assessing EFL teachers' mobile teaching affordances. The data were meticulously scrutinised employing the Mokken package version 3.0.6 (van der Ark et al., 2021) within the R (R Core Team, 2021). This study was conducted by the absence of a valid measure for assessing EFL teachers' attitudes towards mobile teaching affordances and their significance. Initially, a 59-item inventory was developed after a thorough review of related literature. Ten experts were asked to review and provide feedback on the items. As a result, some items were added, deleted or modified based on the reviewers' comments. Subsequently, a 36-item scale was finalised, encompassing four dimensions - assessments connectivity, context-sensitivity, mobility and interaction with others.

The research findings revealed that the Invariant Item Ordering (IIO) assumption was violated for some items in the scale. The IIO assumption is crucial in the Mokken Scale Analysis (MSA) as it determines whether the ordering of items remains consistent across all respondents, regardless of their varying levels of traits. The analysis indicated that the scale did not meet the IIO assumption, implying that ordering of items was not consistent for all respondents, regardless of their trait level. The study utilised the HT coefficient to assess the overall IIO, and the HT value of 0.154 indicated inaccurate item ordering. Consequently, 24 items were found to violate the IIO assumption, with only 10 of these items' violations being significant of IIO, and thus requiring removal. The violations were determined by the number of times an item response function (IRF) intersected with the IRF of other items. For instance, item 2 had 11 violations, but only five of these violations were significant. Similarly, the IRF for item 3 intersected with the IRF of six other items, with only one being significant. The method of backward selection was utilised in order to eliminate items that were in violation of IIO, leading to the exclusion of ten items. Based on the results, a total of ten items (1, 2, 3, 4, 5, 12, 16, 26, 29, and 36) violated the IIO and were removed. Subsequently, the quality of the 26-item scale was analysed, with the total scalability found to be 0.395. All item pairs exhibited positive scalability, with the coefficients of scalability for the items surpassing the predetermined threshold. Furthermore, the findings indicated that although the five remaining items violated the IIO, their violations were not significant, suggesting a good fit of MHM and DMM to the data. The study also analysed the monotonicity assumption, which posits that as individuals' ability levels increase, the likelihood of providing an accurate response or supporting a response option will also increase. The results demonstrated that all items of the scale had met the assumption of monotonicity, supporting the ordering of respondents based on their total scores. Additionally, as a preliminary check, Cronbach's alpha indicated a high degree of reliability and internal consistency with the sample. The outcomes of the correlation between items and the total score also revealed that each item in the scale superbly captured the essence of the identical underlying construct. However, upon validating the data through Mokken scale analysis, ten items were removed. The removal of these items led to improvements in the values of the items and total scalability coefficients, with all item pairs being nonnegative. Furthermore, the assumption of monotonicity was maintained, all items were found to be unidimensional, non-significant violations of IIO were identified, and both MHM and DMM could fit well with the data.

Accordingly, the results of MSA indicated that the inventory possesses construct validity, which is consistent with a similar study by Rostami (2021), who designed and validated an inventory of mobile learning affordances from the learners' perspectives using confirmatory factor analysis. Moreover, Alkhudair (2020) surveyed m-learning and measured the users' attitudes towards it electronically. The reliability of the scale was found to be 0.87, which is considered acceptable. Additionally, Parsons et al. (2016) developed a questionnaire to measure six remarkable capabilities of mobile devices, encompassing the domains of portability, data acquisition, interpersonal connection, interface engagement, contextual and dynamic knowledge acquisition, and even outdoor adaptability.

6. Conclusion

The findings of the investigation suggest that the finalised scale exhibits high reliability as evidenced by the four different reliability coefficients used in the study - the latent class reliability coefficient (LCRC), Mokken scale (MS) reliability or ρ, Cronbach's alpha and Lambad-2. All the values of these coefficients were greater than 0.90, signifying the high reliability of the scale. Such observations imply that the scale exhibits a steadfastness in assessing the planned construct, making it suitable to be confidently used in research and practice. The study also found that all items of the scale met the assumption of monotonicity, supporting the arrangement of participants in terms of their cumulative scores. However, the scale violated the invariant item ordering (IIO) assumption, with 24 items violating IIO, and only 10 of these items having significant violations. To address this, the research employed the technique of backward selection to eliminate the items that infringed upon the principle of IIO, which consequently led to the removal of ten items.

In conclusion, the developed scale can be utilised to assess EFL teachers' attitudes towards mobile teaching affordances and identify areas where teachers need to enhance their knowledge and skills related to mobile teaching affordances. The results of this investigation offer a conceptual framework for teachers to enhance teaching effectively. The study's implications extend to the development of training programmes for EFL teachers to enhance their teaching by becoming more familiar with the affordances of mobile devices. The results of this study also underscore the growing importance of technology in the field of education and its potential to personalise learning, emphasising the need for serious consideration of this issue in educational systems. Additionally, the outcomes of the study have significant implications for forthcoming research that aims to explore the correlation between EFL teachers' attitudes towards the possibilities offered by mobile teaching and their instructional methods. Additionally, it sheds light on the effectiveness of mobile teaching affordances in improving EFL teaching and learning outcomes.

Acknowledgments

The authors wish to convey their profound appreciation to Mrs. Sanaz Samir for her invaluable input and guidance throughout this research. The authors are also deeply obliged to the teachers who graciously dedicated their time and exerted considerable effort to facilitate this investigation.

References

Adler, M., & Brodin, U. (2011). An IRT validation of the Affective Self Rating Scale. *Nordic Journal of Psychiatry, 65*(6), 396–402. https://doi.org/10.3109/08039488.2011.577187

Alaghbary, G. S. (2021). Mobile-assisted language learning in Saudi college education: Exploring learners and teachers' perceptions and practices. *Asian ESP Journal, 17*(3.1), 8–31. Retrieved from https://www.asian-esp-journal.com/esp-march-2020-17-3-1/

Alkhudair, R. Y. (2020). Mobile assisted language learning in Saudi EFL classrooms: Effectiveness, perception, and attitude. *Theory and Practice in Language Studies, 10*(12), 1620–1627. http://dx.doi.org/10.17507/tpls.1012.16

Baghaei, P. (2021). *Mokken scale analysis in language assessment*. Waxmann Verlag.

Bahari, A. & Salimi, M. (2021). Challenges and affordances of developing receptive and productive skills via technology-based interaction. *Computer-Assisted Language Learning Electronic Journal, 22*(1), 22–52.

Chen, C. H., & Tsai, C. C. (2021). In-service teachers' conceptions of mobile technology-integrated instruction: Tendency towards student-centered learning. *Computers & Education, 170*, 104224.https://doi.org/10.1016/j.compedu.2021.104224

Chou, Y. H., Lee, C. P., Liu, C. Y., & Hung, C. I. (2017). Construct validity of the Depression and Somatic Symptoms Scale: Evaluation by Mokken scale analysis. *Neuropsychiatric Disease and Treatment, 13*, 205–211. https://doi.org/10.2147/NDT.S118825

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

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

Egessa, M., Liyala, S., & Ogara, S. (2021). Affordances theory in information and communication technology for development (ICT4D) research. *International Research Journal of Modernization in Engineering Technology and Science*, *3*(1), 324–332. Retrieved from https://www.irjmets.com/uploadedfiles/paper/volume_3/issue_1_january_2021/5621/final/fin_irjmets1628271803.pdf

Firoozi, F. (2021). Mokken scale analysis of the reading comprehension section of the International English Language Testing System (IELTS). *International Journal of Language Testing, 11*(2), 91–108. Retrieved from https://www.ijlt.ir/article_138059.html

García-Martínez, I., Fernández-Batanero, J. M., Cobos Sanchiz, D., & Luque de La Rosa, A. (2019). Using mobile devices for improving learning outcomes and teachers' professionalization. *Sustainability, 11*(24), 6917. https://doi.org/10.3390/su11246917

Gaudry-Perkins, F., & Dawes, L. (2011). M-learning: A powerful tool for addressing MDGs. *MDG Review*, 6–12. https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2012/05/AlcatelLucent_120504.pdf

Gunter, G. A., & Braga, J. D. C. F. (2018). Connecting, swiping, and integrating: Mobile apps affordances and innovation adoption in teacher education and practice. *Educação em Revista, 34,* e189927. https://doi.org/10.1590/0102-4698189927

Gutiérrez-Colón, M., Frumuselu, A. D., & Curell, H. (2023). Mobile-assisted language learning to enhance L2 reading comprehension: A selection of implementation studies between 2012–2017. *Interactive Learning Environments*, *31*(2), 854–862. . https://doi.org/10.1080/10494820.2020.1813179

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255–282. https://doi.org/10.1007/BF02288892

Hoi, V. N., & Mu, G. M. (2021). Perceived teacher support and students' acceptance of mobile-assisted language learning: Evidence from Vietnamese higher education context. *British Journal of Educational Technology*, *52*(2), 879–898. https://doi.org/10.1111/bjet.13044

IBM Corp. (2016). *IBM SPSS statistics for Windows* (Version 24.0) [Computer software]. IBM Corp.

Ibrahim, A. A., & Kadiri, G. C. (2018). Integrating mobile phones in teaching auditory and visual learners in an English classroom. *English Language Teaching, 11*(12), 1–10. https://doi.org/10.5539/elt.v11n12p1

Keskin, N. O., & Metcalf, D. (2011). The current perspectives, theories and practices of mobile learning. *TOJET: The Turkish Online Journal of Educational Technology, 10*(2), 202–208. Retrieved from http://www.tojet.net/articles/v10i2/10220.pdf

Kirschner, P. A., Martens, R. L., & Strijbos, J. W. (2004). CSCL in higher education? A framework for designing multiple collaborative environments. In J. W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), *What we know about CSCL: And implementing it in higher education*. (pp. 3–30). Springer.

Kukulska-Hulme, A. (2012). Mobile-assisted language learning. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3701–3709). Blackwell.

Lee, C. P., Liu, C. Y., & Hung, C. I. (2017). Psychometric evaluation of a 6-item Chinese version of the Hamilton Depression Rating Scale: Mokken scaling and item analysis. *Asia-Pacific Psychiatry*, *9*(3), e12287. https://doi.org/10.1111/appy.12287

Li, R. (2023). Effects of mobile-assisted language learning on EFL learners' listening skill development. *Educational Technology & Society*, *26*(2), 36–49. https://www.jstor.org/stable/48720994

Ligtvoet, R., van der Ark, L. A., te Marvelde, J. M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, *70*(4), 578–595. https://doi.org/10.1177/0013164409355697

MacCallum, K., & Kinshuk. (2006, 22–25 October). *Mobile technology in facilitating learning goals* [Paper presentation]. mLearn 2006: The 5th World Conference on Mobile Learning, Alberta, Canada. https://auspace.athabascau.ca/handle/2149/1236

MacCallum, K., Day, S., Skelton, D., & Verhaart, M. (2017). Mobile affordances and learning theories in supporting and enhancing learning. *International Journal of Mobile and Blended Learning (IJMBL)*, *9*(2), 61–73. https://doi.org/10.4018/IJMBL.2017040104

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods, 9*(3), 354–368. https://doi.org/10.1037/1082-989X.9.3.354

Meijer, R. R., Tendeiro, J. N., & Wanders, R. B. K. (2015). The use of nonparametric item response theory to explore data quality. In S. P. Reise, & D. A. Revick (Eds.), *Handbook of item response theory modelling: Applications to typical performance assessment* (pp. 85–110). Routledge.

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter Mouton. https://doi.org/10.1515/9783110813203

Molenaar, W., & Sijtsma, K. (2000). *MSP5 for Windows user's manual.* Iec ProGAMMA.

Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M., (2004). *NESTA Futurelab Report 11: Literature review in mobile technologies and learning.* Nesta Futurelab.

Nyiri, K. (2002). Towards a philosophy of m-learning. In M. Milrad, U. Hoppe, & Kinshuk (Eds.), *Proceedings of the IEEE international workshop on wireless and mobile technologies in education* (pp. 121–124). IEEE Computer Society. https://doi.org/10.1109/WMTE.2002.1039233

Oxford English Dictionary. (n.d). Affordance. In *OED.com.* Retrieved May 9, 2024, from

Parsons, D., Thomas, H., & Wishart, J. (2016). Exploring mobile affordances in the digital classroom. In I. A. Sánchez, & P. Isaías (Eds.), *Proceedings of the 12th international* *conference mobile learning 2016* (pp. 43–50). International Association for Development of the Information Society (IADIS). https://www.iadisportal.org/digital-library/exploring-mobile-affordances-in-the-digital-classroom

R Core Team. (2021, May). *R: A language and environment for statistical computing* (Version 4.1.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reeves, J. L., Gunter, G. A., & Lacey, C. (2017). Mobile learning in pre-kindergarten: Using student feedback to inform practice. *Journal of Educational Technology & Society, 20*(1), 37–44.

Rostami, Z. (2021). *Development and validation of mobile learning affordance inventory for Iranian EFL context* [Unpublished master's thesis]. Tabaran Institute of Higher Education, Mashhad.

Sharples, M. (2002). Disruptive devices: Mobile technology for conversational learning. *International Journal of Continuing Engineering Education and Life-Long Learning, 12* (5/6), 504–520. https://dx.doi.org/10.1504/IJCEELL.2002.002148

Sijtsma, K., & van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology, 70*(1), 137–158. https://doi.org/10.1111/bmsp.12078

Sijtsma, K., & Molenaar, I. W. (2002*). Introduction to nonparametric item response theory*. SAGE Publications.

Soper, D. S. (2020). *A-priori sample size calculator for student t-tests* [Software]. Free Statistics Calculators. Retrieved from https://www.danielsoper.com/statcalc/calculator.aspx?id=47

Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*, *12*(74), 1–16. https://doi.org/10.1186/1471-2288-12-74

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement*, *74*(5), 809–822. https://doi.org/10.1177/0013164414529793

Tabares, J. V., Butner, J. E., Bryan, C. J., & Harris, J. A. (2021). Mokken scale analysis of lifetime responses on the Columbia Suicide Severity Rating Scale's severity of ideation subscale. *Assessment, 28*(6), 1624–1634. https://doi.org/10.1177/1073191120913626

Tabatabaee-Yazdi, M., Motallebzadeh, K., Baghaei, P. (2021). A Mokken scale analysis of an English reading comprehension test. *International Journal of Language Testing*, *11*(1), 132–143. Retrieved from https://www.ijlt.ir/article_130373.html

Togaibayeva, A., Ramazanova, D., Yessengulova, M., Yergazina, A., Nurlin, A., & Shokanov, R. (2022). Effect of mobile learning on students' satisfaction, perceived usefulness, and academic performance when learning a foreign language. *Frontiers in Education, 7*(2002).https://doi.org/10.3389/feduc.2022.946102

van der Ark, L. A., van der Palm, D. W., & Sijtsma, K. (2011). A latent class approach to estimating test-score reliability. *Applied Psychological Measurement*, *35*(5), 380–392. https://doi.org/10.1177/0146621610392911

van der Ark, L. A., Koopman, L., Straat, J. H., & Van den Bergh, D. (2021, March 9). *Mokken package (*Version 3.0.6) [Computer software]. The Comprehensive R Archive Network. Retrieved from https://cran.r-project.org/package=mokken

van Schuur, W. H. (2011). *Ordinal item response theory: Mokken scale analysis*. SAGE Publications.

Watson, R., Egberink, I. J. L, Kirke, L., Tendeiro, J. N., & Doyle, F. (2018). What are the minimal sample size requirements for Mokken scaling? An empirical example with the Warwick-Edinburgh Mental Well-Being Scale. *Health Psychology and Behavioral Medicine*, *6*(1), 203–213. https://doi.org/10.1080/21642850.2018.1505520

Welsh, K. E., Mauchline, A. L., Powell, V., France, D., Park, J. R., & Whalley, W. B. (2015). Student perceptions of iPads as mobile learning devices for fieldwork. *Journal of Geography in Higher Education*, *39*(3), 450–469. https://doi.org/10.1080/03098265.2015.1066315

Wind, S. A. (2017). An instructional module on Mokken scale analysis. *Educational Measurement: Issues and Practice*, *36*(2), 50–66. https://doi.org/10.1111/emip.12153

Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.

Appendix A

Teachers' Attitudes towards Mobile Teaching Affordances

Dear English Teachers,

The following inventory asks for information regarding your experience with mobile affordances. Please take a few minutes to indicate how strongly you agree or disagree with the following statements by selecting the appropriate answer following the statement.

Thank you for your cooperation.

N	Items	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1. Connectivity → I can/usually use my mobile phone to/for …						
*1	access online libraries.					
*2	distribute course materials to my students.					
*3	check cultural events.					
*4	check the news of the institution.					
*5	check my portal.					
6	check emails.					
7	send homework assignments to the students.					
8	send voice or video files to my students.					
9	send videos or pictures to help the initiation of a discussion.					
10	send educational games to my students.					
11	share my beliefs, ideas, experiences and viewpoints.					
*12	read electronic books.					
13	search information.					
14	collect data using different ways such as Google Docs/Forms.					
15	record or download audio/video clippings and play them inside the classroom.					
*16	send posts/movies to Instagram, YouTube or Facebook.					
2. Context Sensitivity → I can/usually use my mobile phone to/for …						
17	record classes (for students) to review in the future.					
18	take photos of class (students) presentations.					
19	save different types of course content to use in the future.					
20	to take notes in my classes.					
21	to fill questionnaires.					
22	design and develop online questionnaires, exams and tests.					
23	access Learning Management System (LMS).					
24	attend my online classes on different platforms (Adobe Connect, Sky Room, or Big Blue Button) through mobile technology.					

25	use a mobile dictionary to find out the meaning of the words.					
3. Mobility & Interaction with Others → I can/usually use my mobile phone to/for …						
*26	learn/teach with no restriction to a certain time or a specific place.					
27	make contact with my colleagues.					
28	arrange my teaching appointments.					
*29	ask my students questions by sending messages.					
30	get answers to my questions.					
31	enhance communication in my classroom.					
32	enhance collaborative learning in my classroom.					
4. Assessments → I can/usually use my mobile phone to/for …						
33	assess and monitor my students' learning.					
34	take tests and quizzes using Google Docs, etc.					
35	provide students with feedback and error corrections.					
*36	check my students' homework assignments.					

Note. The *-items are deleted from the inventory during the validation process.