BARÜ Bartın University Journal of Faculty of Education BUEFAD
2023, Volume 12, Issue 3, 570-578
dergipark.org.tr/buefad
DOI: 10.14686/buefad.1197656

Investigating the Validity and Reliability of the Mobile Application Rating Scale

Yıldız ÖZAYDIN AYDOĞDU [a], ***Hatice YILDIZ DURAK*** [*b], ***Elif AKGÜN*** [c]

a Instructor, Nevşehir Hacı Bektaş Veli University, https://orcid.org/0000-0002-7433-3057
b Dr., Necmettin Erbakan University, https://orcid.org/0000-0002-5689-1805 *hatyil05@gmail.com
c Bartın University, https://orcid.org/0000-0002-5689-1805

Research Article
Received: 1.11.2022
Revised: 10.4.2023
Accepted: 10.4.2023

Abstract

The purpose of this study is to perform the validity and reliability study of the Mobile Application Rating Scale (MARS) for a mobile learning system that contains e-contents in different forms to improve digital parenting competencies for adults. This scale was prepared by Stoyanov et al. (2015) and adapted into Turkish for the e-pulse application by Korkmaz and Arıkan (2021). In this study, the scale items were rearranged in the context of the study to evaluate "digital parenting". This scale aims to evaluate the application with four sub-dimensions of "participation, functionality, aesthetics, and knowledge", and it has 21 items. The study was carried out with 181 participants. Considering the validity results of the scale, the model with the highest validity among the single-factor model, four-factor model, and second-level confirmatory factor analysis models was the four-factor model, as in the original scale. Factor loads vary between 0.55 and 0.91. Considering the reliability results, the average variance extracted for the participation dimension was 0.69, for functionality 0.78, for aesthetics 0.75, for knowledge dimension 0.63, and the average variance extracted for the whole scale was 0.70. In addition, alpha reliability (stratified alpha reliability) was 0.92 for participation, 0.94 for functionality, 0.92 for aesthetics, 0.92 for knowledge, and 0.97 for the whole scale. In the context of these findings, the validity and reliability values of the scale were high.

Keywords: Mobile learning system, mobile app rating, digital parenting, scale adaptation, validity, reliability

Mobil Uygulama Derecelendirme Ölçeğinin Geçerlik ve Güvenirliğinin İncelenmesi

Öz

Bu çalışmada amaç yetişkinlere yönelik dijital ebeveynlik yeterliklerini geliştirmeye yönelik farklı formlarda e-içerikler barındıran bir mobil öğrenme sistemine yönelik Mobil Uygulama Derecelendirme Ölçeği'nin (MUD) geçerlik ve güvenirlik çalışmasının gerçekleştirilmesidir. Bu ölçek Stoyanov et al. (2015) tarafından geliştirilmiş, Korkmaz ve Arıkan (2021) tarafından e-nabız uygulamasına yönelik olarak Türkçe'ye uyarlama çalışması yapılmıştır. Bu çalışmada ise eğitim alanında geliştirilen "Dijital ebeveynlik" uygulamasının değerlendirilmesi amacıyla ölçek maddeleri çalışma bağlamında yeniden düzenlenmiştir. Bu ölçek eğitsel bir mobil uygulamanın kullanıcıları tarafından katılım, işlevsellik, estetik ve bilgi dört alt boyutunu ve 21 madde çerçevesinde değerlendirilmesini amaçlamaktadır. Çalışma 181 katılımcıyla gerçekleştirilmiştir. Ölçeğin geçerlik sonuçlarına bakıldığında; tek faktörlü model, dört faktörlü model ve ikinci düzey doğrulayıcı faktör analizi modellerinden en yüksek geçerliliğe sahip model, ölçeğin orijinalinde de olduğu gibi dört faktörlü model olmuştur. Faktör yükleri 0.55 ve 0.91 arasında değişiklik göstermektedir. Güvenirlik sonuçlarına bakıldığında ise ortalama açıklanan varyansın katılım boyutu için 0.69, işlevsellik için 0.78, estetik için 0.75, bilgi boyutu için 0.63 olduğu, ölçeğin tamamının ortalama açıklanan varyansının ise 0.70 olduğu görülmüştür. Ayrıca alfa güvenilirliği (tabakalı alfa güvenirliği) katılım için 0.92, işlevsellik için 0.94, estetik için 0.92, bilgi için 0.92 ve ölçeğin tamamı için 0.97 olduğu görülmektedir. Bu bulgular bağlamında ölçeğin geçerlik ve güvenirlik değerlerinin yüksek olduğu sonucuna ulaşılmıştır.

Anahtar kelimeler: Mobil öğrenme sistemi, mobil uygulama derecelendirme, dijital ebeveynlik, ölçek uyarlama, geçerlilik, güvenilirlik.

To cite this article in APA Style:

Özaydın Aydoğdu, Y. & Yıldız Durak, H. & Akgün, E. (2023). Investigating the validity and reliability of the mobile application rating scale. *Bartın University Journal of Faculty of Education, 12*(3), 570-578.. https://doi.org/10.14686/buefad.1197656

INTRODUCTION

Today, technology is developing rapidly and the use of mobile devices is also becoming widespread (Yildiz-Durak, 2018). This situation also increases the number of mobile applications to a great extent. The prevalence of mobile devices causes desktop computer applications to be replaced by mobile applications. Mobile applications create a dynamic change by increasing the workspace, environment, and mobility of users (Costa, 2003). In this way, it is possible to use mobile applications with the opportunity to access them anytime and anywhere. Since it is foreseen that the use of these systems will continue to increase in the future, the quality of mobile applications is of great importance for users (Şimşek, Betin-Can, & Can, 2016; Yildiz-Durak, 2019a, b). While mobile devices are typically used to make phone calls, the transition to smartphones is happening rapidly with the internet. Thanks to the developments in the hardware and software of mobile devices, the installation of mobile applications allows mobile phones to offer new services (Aloul, Zahidi, & El-Hajj, 2009). In addition to allowing users to access various applications, mobile devices have limitations such as device screen size, resolution, and different input features (Griggs, Bridges, & Rempel, 2009). For this reason, it is important to pay attention to interface designs while preparing applications for mobile devices.

Since mobile internet devices are personal compared to fixed internet devices, user identity is more important. However, mobile systems have many inherent limitations when developing content and user interfaces. Mobility restrictions include restrictions on user, device, communication, and technology. Communication and device restrictions include high error rates, disconnections, high noise, low memory and processing capacity, small screen size, and limited battery power resources (Subramanya, & Yi, 2006).

The developments and changes experienced with the increasing prevalence of mobile systems in the world reveal the concept of mobile learning (Yildiz-Durak, Canbazoğlu-Bilici, & Baran, 2023). Therefore, mobile systems also affect educational environments. Educational environments have always been in a relationship with the current technological tools (Yildiz-Durak, 2019a). Today, with the widespread use of mobile devices such as mobile phones and tablets, this technology change necessitates the transmission of educational content via mobile devices (Çakır & Arslan, 2013). Mobile devices have features such as being portable, running multimedia, and working with the internet, and their development over time makes them attractive to use in education (Yildiz-Durak, Canbazoğlu-Bilici, & Baran, 2023).

When the studies on mobile application evaluation are examined, Stoyanov, Hides, Kavanagh, Zelenko, Tjondronegoro, & Mani, (2015) developed a mobile application rating scale (MARS) in English to evaluate different mobile applications in the field of health. Korkmaz and Arıkan (2021) carried out this scale by adapting a mobile application in the field of health for students to use in their research. For MARS healthcare, German (Messner et al., 2020), Arabic (Bardus et al., 2020), Japanese (Yamamoto et al., 2022), Korean (Hee Ko, Kim, Lee, Lee, & Stoyanov, 2022) and Turkish (Korkmaz & Arıkan, 2021) adapted this scale into their languages and their validity and reliability were evaluated. When the literature is examined, adaptation, validity, and reliability analysis have been made for the MARS health field.

Stoyanov et al. (2015), the mobile application rating scale was developed to determine the quality of the application. Within the scope of this purpose, the scale includes participation, functionality, aesthetics and knowledge sub-dimensions. Participation dimension enables users to determine whether this application is fun, interesting, customizable, interactive and suitable for the target audience as a result of users' interaction with the application. Functionality dimension enables the determination of the performance of the application, that is, whether it works or not, whether it provides ease of use, whether it provides fluency during navigation and whether there is consistency between menus during navigation. The Aesthetic sub-dimension determines whether the placement of the components in the application provides convenience, the quality of the graphics and whether the application graphics are visually attractive. Finally, the information dimension provides information about whether the application contains information suitable for its purpose, the adequacy of the information it covers, the adequacy of the application and the explanations in its content, and whether the information and the source from which the information is obtained are reliable. The creation of these sub-dimensions and their scopes was determined by the literature review carried out in the original study. As a result of this screening, the sub-dimensions of the scale were determined with three studies that were comprehensive in usability.

In this study, the validation study of the scale developed by Stoyanov, Hides, Kavanagh, Zelenko, Tjondronegoro, & Mani, (2015) and adapted into Turkish by Korkmaz and Arıkan (2021). The scale adapted to Turkish by Korkmaz and Arıkan (2021) for the e-pulse application is discussed for the mobile learning system in the context of this study. In this study, to evaluate the “digital parenting” application developed in the field of

education, the scale items were rearranged in the context of the study, and it was aimed to determine the validity and reliability of the scale. This scale aims to evaluate an educational mobile application by users within the framework of participation, functionality, aesthetics, and knowledge dimensions.

METHOD

This research is a scale adaptation study. The validity and reliability study of the Turkish version of the Mobile Application Rating Scale (MARS) was carried out. The MARS scale developed by Stoyanov et al. (2015) was translated into Turkish by Korkmaz and Arıkan (2021). Stoyanov et al. (2015), a pilot application was carried out on 9 applications used in the field of health, and data were collected for the reliability tests of the scale developed over 50 application users used in the field of health. The original scale is in 5-point Likert type and consists of 6 sections. The first section consists of 5 questions about entertainment (e.g "Is the app fun/entertaining to use? Does it have components that make it more fun than other similar apps?"), interest, customization, interactivity and target group under the engagement dimension. The second section consists of 4 items related to performance (e.g "How accurately/fast do the app features (functions) and components (buttons/menus) work?"), ease of use, navigation and gestural design under the functionality dimension. The third section consists of 3 items under the titles of layout (e.g "Is arrangement and size of buttons, icons, menus and content on the screen appropriate?"), graphics and visual appeal related to the aesthetic dimension. The fourth section consists of 4 items related to the quality of the information (e.g "Is app content correct, well written, and relevant to the goal/topic of the app?"), quantity of the information, visual informations and credibility of the source under the information dimension. In addition, there are 4 questions to determine the subjective quality (e.g "Would you recommend this app to people who might benefit from it?") of the application under the fifth section of the original scale. Finally, 5-point Likert-type questions (e.g "This app has increased my awareness of the importance of addressing the health behaviour"), and an open-ended question "what you want to add" were added to the scale in order to determine the perceived effect in the sixth section. Since the Turkish version of the scale was prepared in advance, the study was carried out only to determine its psychometric properties. The MARS scale, which has a Turkish form, consists of four sub-dimensions and 21 items: "participation (e.g. "The e-nabız application is fun to use.")", "aesthetics (e.g. "All components (like swiping) between menus in the app are consistent across screens.")", "functionality (e.g. "Learning how to use the e-nabız app is easy.")", and "knowledge (e.g. "The content of the e-nabız application is related to the subject of the application.")". In the study conducted by Korkmaz and Arıkan (2021), the first 4 sections in the original scale were discussed, and the sections where the subjective quality of the application and the perceptible effect were measured were not included in the adaptation. Likewise, in this study, a study in the field of educational sciences was carried out on 21 items used by Korkmaz and Arıkan (2021) to adapt this scale in the field education. For adaptation, this expression has been changed to "dijital ebeveynlik" in all items with the name "e-nabız" in the scale. Validity and reliability studies for the mobile application rating scale were carried out using the "Digital parenting" application. This application was created in 5 different types in the system using the demographic information of the people, parental mediation scale scores, digital parenting self-efficacy scale scores and browsing data (commenting, rating, content viewing time, number of sharing, adding to playlist status, etc.). It is a system that makes suggestions to users among 141 multimedia environments. This study was carried out with individuals whose navigation data were included on this system and analyzes were made for the validity and reliability of the scale with the obtained data. For the validity studies of the scale, a single factor, four factor and second-level confirmatory factor analysis model was used and model-data fit values were calculated. For the reliability studies of the scale, the mean variance and descriptive values of the sub-dimensions and the relationship between these sub-dimensions were calculated. The results obtained are included in the findings section.

Data Collection

During the implementation process of the study, the measurement tool was delivered to the participants in the online environment via "Google Forms". It was shared with 181 parents who used the "digital parenting" application and aged 18-64 adult of the measurement tool. However, since not all parents using the "digital parenting" application could be reached, participants were selected through appropriate sampling.

Research Ethic

The study was carried out within the scope of the project named "Developing a Suggestion System for Increasing Parents' Digital Parenting Competencies ". Ethics committee approval numbered 2020-SBB-0278 was obtained for this project and this permission is included in the appendices.

FINDINGS

After arranging the obtained data in the R program, the randomness of the missing data was examined with the naniar (Tierney, 2021) package. The results obtained from Little's analysis of the randomness of missing data ($\chi^2 = 2269.526$, df = 13535, p = 1) and the missing data pattern created with the mice (van Buuren & Groothuis-Oudshoorn, 2011) package showed that the missing data were randomly distributed. In the missing data pattern data sets in Figure 1, the missing data are completely randomly distributed.

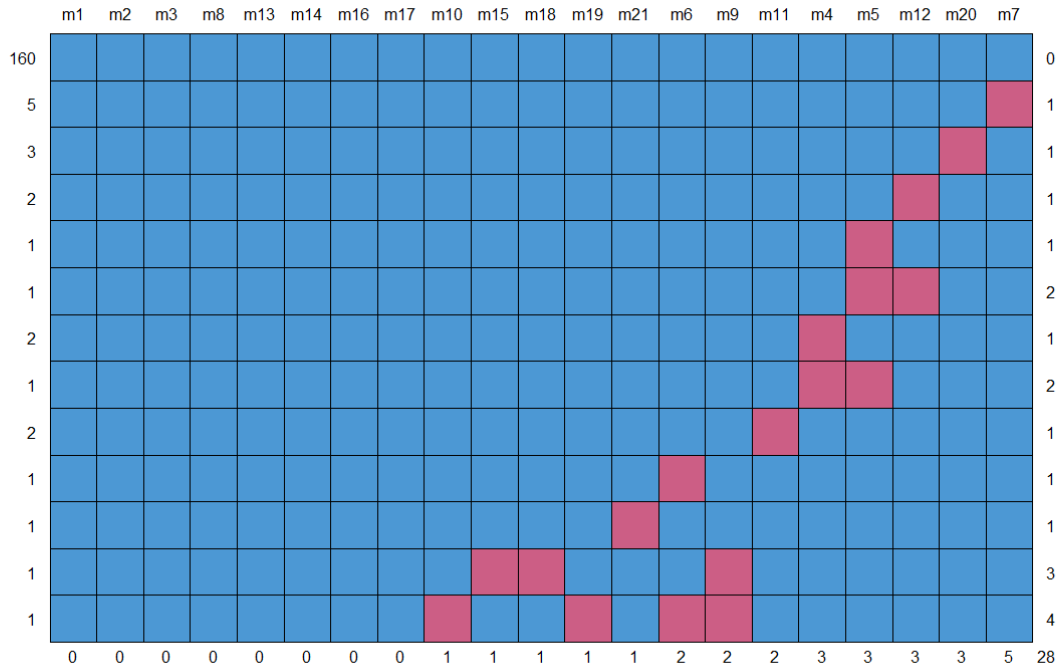

Figure 1. *Distribution of missing data*

Looking at the first line of Figure 1, the data of 160 individuals out of 181 individuals are complete. In the rest of the figure, the numbers in the left margin indicate the number of missing data observed in individuals, while the numbers in the right margin show how many items are missing in the same individuals. The numbers in the bottom margin indicate the number of missing data in the items. When this structure is examined, the variables do not form a deficiency in a certain region depending on each other, that is, they are completely randomly distributed. Based on this randomness, the missing data were assigned with the predictive mean matching method. With the obtained data set, the psychometric properties of the scale in the adaptation process were tested.

The measurement models tested in the original MARS scale were also examined and compared in this study. For this purpose, i) the single-factor model, ii) the four-factor model, and iii) the second-level confirmatory factor analysis model were created by using a robust maximum likelihood method. In the model creation process, lavaan (Rosseel, 2012) and semPlot (Epskamp, 2022) packages were used. Diagrams of the models are given in Figures 2, 3, and 4.

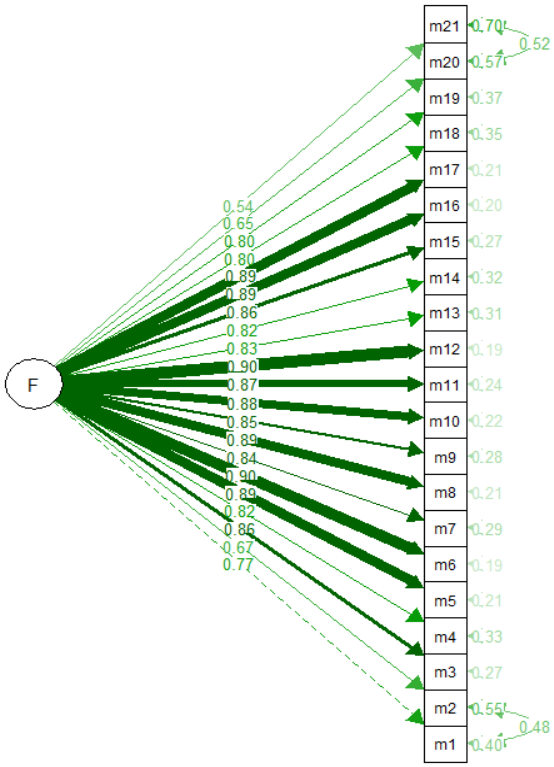

Figure 2. *Single factor model (Model 1a)*

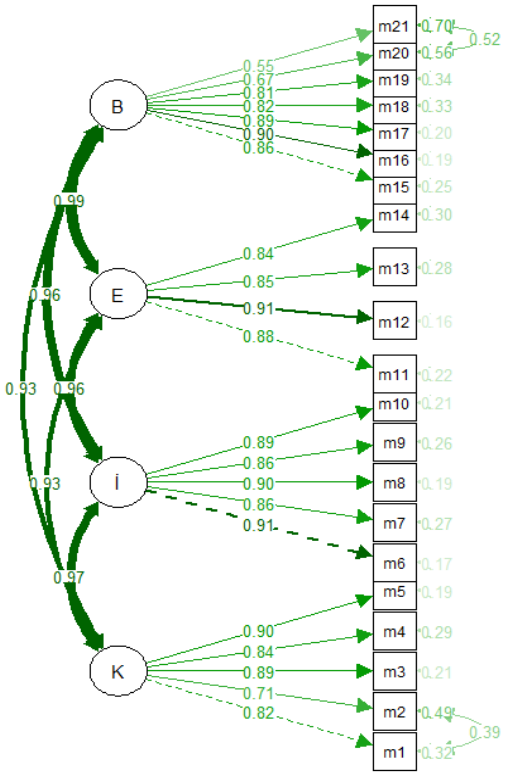

Figure 3. *Four-factor model (Model 2a)*

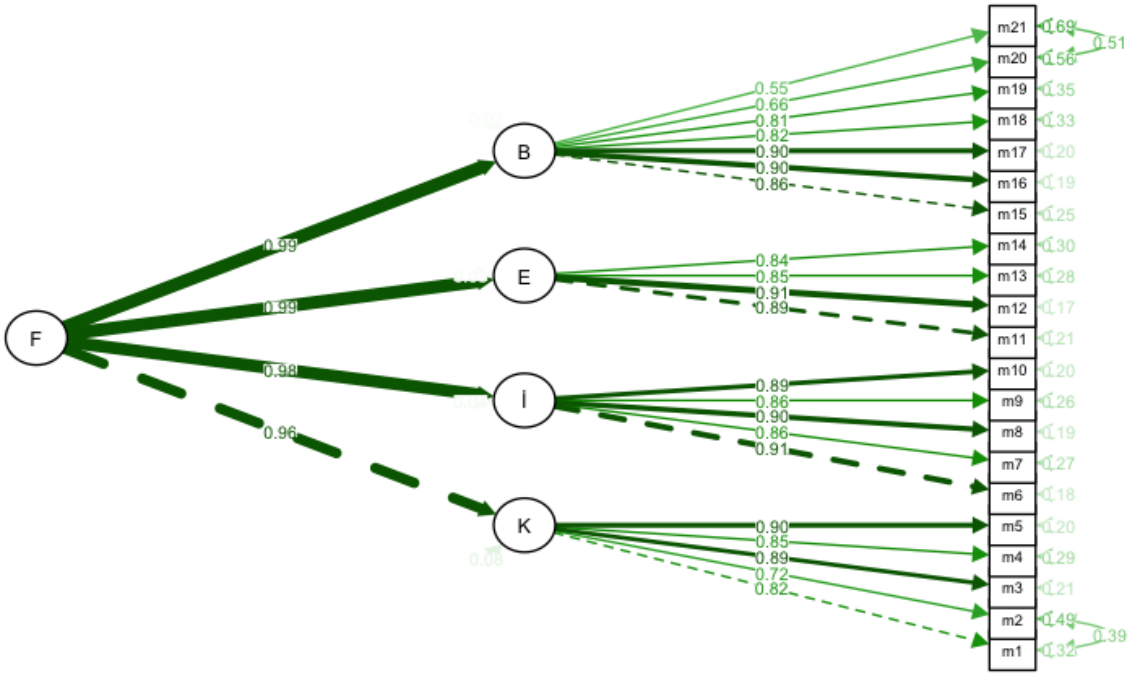

Figure 4. *Second-level confirmatory factor analysis model (Model 3a)*

When the diagrams in Figures 2, 3, and 4 were examined, correlations between m1 and m2 and m20 and m21 items were added considering the modification indices in each model. "The digital parenting app is fun to use. (m1)" and "The use of digital parenting application is interesting. (m2)" items were examined, and it was thought that the fun application would also increase the interest of the participants. Therefore, parts of these items that are not explained by the underlying factor may be due to a structure or structures that are different and similar to the structure examined. Similarly, "Digital parenting app comes from a trusted source. (m20)" and "Digital parenting application is academically supported. The variances of the "(m21)" items, which are not explained by the factor, may originate from other similar structures or structures. Because in our culture, academia is seen as a reliable source. Models in which correlations were not included among the relevant items were expressed as "Model X", while models in which correlations were added were expressed as "Model Xa". The fit values of the models are given in Table 1.

Table 1. Model-Data Fit Values of the Mobile Application Rating Scale

Model	AIC	BIC	RMSEA	SRMR	TLI	CFI
Model 1	7994.985	8129.322	0.097	0.045	0.916	0.924
Model 1a	7896.061	8036.795	0.081	0.037	0.941	0.947
Model 2	7903.856	8057.384	0.082	0.041	0.938	0.946
Model 2a	7827.081	7987.005	0.068	0.033	0.958	0.964
Model 3	7923.458	8070.589	0.086	0.042	0.933	0.941
Model 3a	7848.397	8001.925	0.073	0.035	0.952	0.958

When the fit values of the models in Table 1 are examined, the data shows the best fit with model 2a. In line with these results, the study continued with the four-factor model (Model 2a). Fit indices of Model 2a reveal evidence of construct validity. When the diagram of the model (Figure 3) is examined, the factor loads are between 0.55 and 0.91. In addition, as seen in Table 2, the average variance extracted (AVE) values for each dimension greater than 0.50 can be shown as proof of convergent validity (Fornell & Larcker, 1981). As another proof of convergent validity, each of the structural reliability values is greater than the AVE values (Fornell & Larcker, 1981).

Table 2. Reliability And Average Variance Extracted Values Of MARS Scale Data (Average Variance Extracted -AVE)

Dimensions	Number of items	Alpha Coefficient/ Stratified Alpha Coefficients	Structural reliability	Average Variance Extracted
Participation	5	0.927	0.920	0.697
Functionality	5	0.947	0.947	0.782
Aesthetic	4	0.927	0.926	0.758
Knowledge	7	0.921	0.922	0.632
MARS Scale	21	0.979	0.981	0.707

Reliability analysis, descriptive statistics, and correlation values of the data obtained from the MARS Scale were calculated using the sort (Robitzsch, 2021), base (R Core Team, 2021), and Hmisc (Harrell, 2021) packages, respectively. Table 2 also includes alpha and stratified alpha coefficients for the reliability of the data obtained from the measurement tool. Since the structure of the measurement tool was confirmed in a four-factor structure, Cronbach's alpha was calculated as the internal consistency coefficient for the sub-dimensions of the scale, while stratified alpha for the reliability of the whole scale was calculated. The obtained values give similar results to the structural reliability values.

Table 3. The Descriptive Values of the MARS Scale Sub-Dimensions and the Correlation Between the Sub-Dimensions

Dimensions	$\underline{X}$	Sd	Participation	Functionality	Aesthetic	Knowledge
Participation	19.64	5.12	.84	.88*	.83*	.81*
Functionality	20.29	5.13		.88	.89*	.87*
Aesthetic	16.55	3.92			.87	.89*
Knowledge	28.61	6.98				.79

* $p<0.05$, Italic values indicate the square root of AVE.

The square root of the AVE for the factors in Table 3 related to the divergence validity of the measurement tool should be greater than 0.50 and the correlation of the relevant factor with other factors (Fornell & Larcker, 1981). The square roots of the AVE of the factors are greater than 0.50 and the correlations are equal or smaller. No strong evidence of divergent validity was obtained.

DISCUSSION & CONCLUSION

In this study, validity and reliability tests were made for the scale developed by Stoyanov, Hides, Kavanagh, Zelenko, Tjondronegoro, and Mani (2015) to evaluate a mobile application in the field of health and adapted into Turkish by Korkmaz & Arıkan (2021) for the e-pulse application. This scale aims to evaluate a mobile application by users within the framework of participation, functionality, aesthetics, and knowledge dimensions. The scale consists of 21 items. The scale was applied by the researchers to 181 participants using the "digital parenting" application via Google forms. The collected data were analyzed in terms of validity and reliability studies. Considering the validity results of the scale, the model with the highest validity among the single-factor model, four-factor model, and second-level confirmatory factor analysis models was the four-factor model, as in the original scale. Factor loads vary between 0.55 and 0.91. Additionally, when the data model fit is examined, the highest agreement in these values was obtained from the four-factor model (TLI=0.958). Considering the reliability results, the average variance extracted for the participation dimension was 0.69, for functionality 0.78, for aesthetics 0.75, for knowledge dimension 0.63, and the average variance extracted for the whole scale was 0.70. When the alpha reliability (stratified alpha reliability) results of the scale are examined, these results are 0.92 for participation, 0.94 for functionality, 0.92 for aesthetics, 0.92 for knowledge, and 0.97 for the whole scale. Analyzes were made for the divergent validity of the scale, but strong evidence for the divergent validity could not be reached in the analysis results. When these results are to be interpreted, the validity and reliability results of the scale are high.

When the results of the MARS scale validity study conducted by Terhorst, Philippi, Sander, Schultchen, Paganini, Bardus, and Messner, (2020) are examined, as a result of the analyzes made, the scale consists of 4 sub-factors, as in this study, and a total of 19 substance has been specified. Item distribution according to factors; There are five items in the participation dimension, four items in the functionality dimension, three items in the aesthetic dimension, and seven items in the knowledge dimension. In the original scale, there are a total of 19 items within the framework of these factors. Korkmaz and Arıkan (2021), in their adaptation study, wrote two questions each for the ease-of-use item under the functionality dimension and the items for placement under the aesthetic

dimension to be more descriptive. As a result of the analyzes made by Korkmaz and Arıkan (2021), it was seen that these new items were again under the factors they belonged to. As a result of the research, it was found that the scale consisted of 21 items in total, including five items in participation, five items in functionality, four items in aesthetics, and seven items in knowledge. In this study, as a result of the analyzes, the item distributions and numbers seen in the study of Korkmaz and Arıkan (2021) were reached. It is seen that the original scale consists of 19 items in the German, Arabic, Japanese and Korean versions under the first four factors.

Terhorst, Philippi, Sander, Schultchen, Paganini, Bardus, ... & Messner, (2020) concluded that this scale is well-structured in terms of evaluating mobile applications developed in the field, in line with the validity and reliability results of the original scale in the field of health. Since the dimensions of participation, functionality, aesthetics, and knowledge in the scale are not specific to the health field but can predict user satisfaction for mobile applications in different areas (e.g. Medicine, Health, Education, books and reference works, entertainment, lifestyle, simulation, social media (Narrillos-Moraza, Gómez-Martínez-Sagrera, Amor-García, Escudero-Vilaplana, Collado-Borrell, Villanueva-Bueno,... & Sanjurjo-Sáez, (2022)), this scale is a well-structured scale for the evaluation of the applications to be developed for educational purposes.

Researchers' Contribution Rate

The contributions of the authors to this study are detailed in Table 4.

Table 4. Researchers' Contribution Rate

Authors	Literature review	Method	Data Collection	Data Analysis	Results	Conclusion
Author 1	☒	☒	☒	☒	☐	☐
Author 2	☐	☒	☐	☐	☒	☒
Author 3	☒	☐	☐	☐	☐	☒

Acknowledgment

This study was carried out within the scope of the TUBİTAK 1001 project numbered 120K927, and entitled "Developing a Suggestion System for Increasing Parents' Digital Parenting Competencies"

REFERENCES

Aloul, F., Zahidi, S., & El-Hajj, W. (2009). Multi factor authentication using mobile phones. International Journal of Mathematics and Computer Science, 4(2), 65-80.

Bardus, M., Awada, N., Ghandour, L. A., Fares, E.-J., Gherbal, T., Al-Zanati, T., & Stoyanov, S. R. (2020). The Arabic version of the Mobile App Rating Scale: development and validation study. *JMIR mHealth and uHealth, 8(3),* e16956.

Costa, P. D. (2003). Towards a services platform for context-aware applications. *Telem-atics, University of Twente, Enschede. Netherlands. August*.

Çakır, H., & Arslan, İ. (2013). Development of course content package for mobile devices. *Bilişim Teknolojileri Dergisi,* 6(3), 24-33.

Durak, H. Y. (2019). Investigation of nomophobia and smartphone addiction predictors among adolescents in Turkey: Demographic variables and academic performance. *The Social Science Journal, 56*(4), 492-517.

Durak, H. Y., Bilici, S. C., & Baran, E. (2023). Engineering Design-Based Arduino Activities in STEM Education. In *Theoretical and Practical Teaching Strategies for K-12 Science Education in the Digital Age* (pp. 66-78). IGI Global.

Epskamp, S. (2022). semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output. R package version 1.1.5. https://CRAN.R-project.org/package=semPlot

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research, 18*(1), 39-50.

Griggs, K., Bridges, L. M., & Rempel, H. G. (2009). Library/mobile: tips on designing and developing mobile web sites. *Code4lib journal*(8).

Harrell, F.E. (2021). Hmisc: Harrell Miscellaneous. R package version 4.6-0. *https://CRAN.R-project.org/package=Hmisc*

Hee Ko, K. K., Kim, S. K., Lee, Y., Lee, J. Y., & Stoyanov, S. R. (2022). Validation of a Korean version of mobile app rating scale (MARS) for apps targeting disease management. *Health Informatics Journal, 28*(2), 14604582221091975.

Korkmaz, S., & Arıkan, G. (2021). A New Tool Used to Evaluate e-Nabız Application: Mobile Application Rating Scale. *Ankara Hacı Bayram Veli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 23*(3), 625-636.

Mendi, O., Sari, M. K., Stoyanov, S., & Mendi, B. (2022). Development and validation of the Turkish version of the Mobile App Rating Scale–MARS-TR. *International Journal of Medical Informatics,* 166, 104843.

Messner, E.-M., Terhorst, Y., Barke, A., Baumeister, H., Stoyanov, S., Hides, L., Kavanagh, D., Pryss, R., Sander, L., & Probst, T. (2020). The German version of the Mobile App Rating Scale (MARS-G): development and validation study. *JMIR mHealth and uHealth, 8*(3), e14479.

Narrillos-Moraza, Á., Gómez-Martínez-Sagrera, P., Amor-García, M. Á., Escudero-Vilaplana, V., Collado-Borrell, R., Villanueva-Bueno, C., Gómez-Centurión, I., Herranz-Alonso, A., & Sanjurjo-Sáez, M. (2022). Mobile apps for hematological conditions: review and content analysis using the mobile app rating scale. *JMIR mHealth and uHealth, 10*(2), e32826.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org/.*

Raio, C. M., Harp, N. R., Brown, C. C., & Neta, M. (2021). Reappraisal—but not suppression—Tendencies determine negativity bias after laboratory and real-world stress exposure. *Affective Science, 2*(4), 455-467.

Robitzsch, A. (2020). sirt: Supplementary item response theory models. R package version 3.9-4. *[Computer Software]. Retrieved from https://CRAN. R-project. org/package= sirt.*

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Simsek, M. V., Betin Can, A., & Can, B. (2016). Android Uygulamalar Için Yazilim Kalite Modeli. *https://open.metu.edu.tr/handle/11511/75286.*

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth, 3*(1), e3422.

Subramanya, S. R., & Yi, B. K. (2006). User interfaces for mobile content. *Computer, 39*(4), 85-87.

Terhorst, Y., Philippi, P., Sander, L. B., Schultchen, D., Paganini, S., Bardus, M., ... & Messner, E. M. (2020). Validation of the mobile application rating scale (MARS). *Plos one, 15*(11), e0241480.

Tierney, N., Cook, D., McBain, M., & Fay, C. (2021). naniar: Data Structures, Summaries, and Visualisations for Missing Data. R package version 0.6.1. *https://CRAN.R-project.org/package=naniar.*

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1-67. DOI 10.18637/jss.v045.i03

Yamamoto, K., Ito, M., Sakata, M., Koizumi, S., Hashisako, M., Sato, M., Stoyanov, S. R., & Furukawa, T. A. (2022). Japanese version of the mobile app rating scale (MARS): development and validation. *JMIR mHealth and uHealth, 10*(4), e33725.

Yildiz Durak, H. (2019). Examining the acceptance and use of online social networks by preservice teachers within the context of unified theory of acceptance and use of technology model. *Journal of Computing in Higher Education, 31*(1), 173-209.

Yildiz-Durak, H. (2018). What Would You Do Without Your Smartphone? Adolescents' Social Media Usage, Locus of Control, and Loneliness as a Predictor of Nomophobia. *Addicta: The Turkish Journal on Addictions*, *5*(3), 543–557. doi:10.15805/addicta.2018.5.3.0025.