End users’ rating of a mHealth app prototype for paediatric speech pathology clinical assessment

Lamya Alabdulkarim

Department of Rehabilitation Sciences, College of Applied Medical Sciences, King Saud University, P.O. Box 10219, Riyadh 11433, Saudi Arabia

ARTICLE INFO

Article history:
Received 15 March 2021
Revised 25 March 2021
Accepted 17 April 2021
Available online 24 April 2021

Keywords:
mHealth
Clinical assessment
Paediatric
Speech pathology

ABSTRACT

Background: Previous studies reported the efficacy of the implementation of information technology in clinical evaluation. No research has addressed the development of mobile applications for the clinical evaluation and diagnosis of paediatric language disorders.

Purpose: This study aims to investigate the usability of a clinical assessment mobile application (app) prototype, the “Paediatric Arabic Language Test” (PALT), to diagnose language disorders among paediatric patients.

Methods: Using the Lewis Computer Software Usability Questionnaire (CSUQ) and a 5-point Likert Scale, data was collected and scored on the usability of the app prototype developed on two mobile platforms- iPhone and iPad- across a general operating system, iOS. A sample of 77 potential end-users rated the usability of the app prototype that they used between 2017 and 2019.

Results: The average CSUQ rating for the app prototype was 75.68 and 53.2% strongly agreed that the information and its organization was clear and easy to understand; 75% of the end users were very satisfied (p < 0.0001). Out of the total items, 68% of studied items on the scale loaded on factor 1 even after subjected to three standardization methods.

Conclusion: The prototype design was judged to be usable. Users reported an effective user interface that allows effective operation. Differences in the factor loading may be explained by cultural factors, type of task and field context.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Mobile applications (apps) have for some time been integrated in health care and have become growing key resource facilitating health care service delivery. This impact has grown even faster during COVID-19 pandemic due to its inherent accessibility. The number of downloadable health related apps (mHealth apps) on Google Play is 43,285 and on Apple App Store is 45,478 for iOS, with an estimated mobile health market of 37 billion US dollars (Puri-Mirza, 2021). However, mobile health apps developers focus on communication and medication and not diagnosis, which is of equal importance in clinical procedures in health care. Recent recommendations indicate that, to enhance diagnosis and health care, digital technology could play a supportive role (Fact sheets, 2020). A gap in this area, as evidenced by a literature review in the field, is therefore of some concern.

Communication disorders are a notable health concern, with a prevalence rate ranging from 5% to 10% worldwide and expected to increase in upcoming years (Overby and Baft-Neff, 2017). Immediate referrals to a speech therapist is necessary to reduce acute communication impairment and limit long-term impacts, including decreased quality of life, reduced academic and vocational opportunities, chronic behavioural, social, and emotional problems (Taylor, Armfield, Dodrill, and Smith 2014).

In speech pathology, most of previous studies on the application of mHealth have focused on home care management and intervention. Systematic reviews of thousands of apps on Google Play and Apple App store have reported few of good quality (Furlong, Morris, Serry, and Erickson 2018), which is of equal importance in clinical procedures in health care. Recent recommendations indicate that, to enhance diagnosis and health care, digital technology could play a supportive role (Fact sheets, 2020). A gap in this area, as evidenced by a literature review in the field, is therefore of some concern.

Communication disorders are a notable health concern, with a prevalence rate ranging from 5% to 10% worldwide and expected to increase in upcoming years (Overby and Baft-Neff, 2017). Immediate referrals to a speech therapist is necessary to reduce acute communication impairment and limit long-term impacts, including decreased quality of life, reduced academic and vocational opportunities, chronic behavioural, social, and emotional problems (Taylor, Armfield, Dodrill, and Smith 2014).
apps were found, observations on apps e-stores revealed most apps focus on telehealth consultations, booking appointments and patient's record viewing.

To date there is no implementation of apps in the diagnosis of speech and language pathology in Saudi Arabia and regionally. Diagnostic sessions in speech pathology, especially language disorders, is a highly specialized and time and effort consuming process. Enhancing delivery of diagnosis allows clinicians to obtain an evidence-based clinical profile of a child's communication disorders, thereby providing valid information for planning interventions (Edwards, and Dukhovny 2017).

Involvement of potential users’ experience in health apps design has been a gold standard in mHealth which play a significant factor in apps ease of usability (van de Kar, and Hengst 2009). It has been established that applying user-centered design during the prototype phase has an impact on increasing its usability and hence return on investment (Marcus, 2005). The aim of this study was to validate the usability of a mHealth app prototype in clinical assessment of language disorders for paediatric population using Lewis CSUQ.

2. Materials and methods

To investigate end users’ rating of usability, this study employed the Computer System Usability Questionnaire (CSUQ) developed for field testing (Lewis, 1995).

2.1. App prototype design overview

On the basis of a review of the literature and international commercial grade standardized clinical assessments, a clinical language assessment framework was developed, the Paediatric Arabic Language Test (PALT). Using a relational database FileMaker Pro®, an app prototype was created to be easy to use in design and function (Claris FileMaker Pro). Fig. 1 displays an overview of PALT design. The prototype included major functional components with multimedia content. Each page had an interface design that enables the user to access information. The interface design was consistent across items. The prototype had 8 sub-tests and a total of 97 culturally and developmentally appropriate items. A manual was compiled for the standard test protocol. Data were collected for two kinds of mobile platforms—iPhone and iPad—across the general class of operating systems (iOS) to facilitate graphic display of the screen and for its excellent user interface after the download and launch of the app prototype.

Fig. 1. Overview of the PALT app prototype.

For item construct relevance to the proposed objectives of the test and the prototype design, face and content validity was reviewed and validated by an expert researcher, two clinicians, and a cohort of 108 student clinicians. The PALT prototype was uploaded and made available via Dropbox™, to be downloaded by end-users.

2.2. Study design

This was a retrospective, cross-sectional study. End-users were recruited from the speech pathology and audiology program at the Department of Rehabilitation Sciences, King Saud University, and from speech pathology clinics in Riyadh city. Initially, end-users received a short orientation session on how to download and use the prototype. Each end user used the prototype for a period of 8 weeks. End-users then tested 583 Saudi children (274 females and 260 males) between 3 and 8 years of age at governmental and private pre-school and school settings as well as outpatient clinics. Two trained co-examiners tested each child at her school setting as shown in Fig. 2. Data was collected with 100% supervision by an American Speech and Hearing Association (ASHA) licensed experienced bilingual speech-language pathology faculty. Each child participated voluntarily and could withdraw at any time. Prior to data collection, the study was approved by King Saud University Research Ethics Committee and reviewed and approved by the Administration and Planning Department at the Ministry of Education. Consent for children to participate was obtained from parents. Field clinical assessment occurred where service delivery is usually administered to this population. The average testing ses-
The session period was 45 min. At the end of testing, end-users rated the usability of the prototype using the CSUQ. Approximately 89% of participants used series Xs, XR and 11 iPhones. The remaining 11% had an iPad 9–7 or 10–2.

2.3. The instrument

Computer System Usability Questionnaire (CSUQ) is a standardized, generalizable and widely applied usability questionnaire to assess users’ satisfaction with system usability in field research (Lewis, 1995). Responses are rated on a 7-point scale from 1 to 7, with 1 corresponding to “strongly disagree” and 7 corresponding to “strongly agree.” In our study however, we applied a classical 5-point Likert scale for the following reasons:

- The population sample is familiar with the 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree)
- Literature suggests that five-point scale appears to be less confusing and to increase response rate and response quality (Buttle, 1996).
- Previous studies found that a 5-point scale is easier to read and more readily comprehensible to respondents (Dawes, 2008; Krosnick, and Presser, 2010)

2.4. Data analysis

Data were compiled and tested using IBM SPSS v22.0. Statistical analysis was done using two-tailed tests and an alpha error of 0.05. A confidence interval of 95% (p-value ≤ 0.05) was set to determine statistical significance. Frequencies and percentages were used to describe the distribution of items and scale. To determine the factors and their related items, an exploratory factor analysis (EFA) using principal components analysis (PCA) with Varimax rotation was run. The selection threshold for the factor loading to be considered significant was set at 0.60 as recommended for this sample size (Hair, Babin, Anderson, and Black 2018). This sample size was sufficient to conduct factor analysis which was more than the recommend minimum (Lewis, 2014; Fischer, and Milfont, 2010).

3. Results

A total of 77 end-users completed the questionnaire. General descriptive data on participants is listed below (Table 1). Fig. 3 displays a normal distribution of the total scores on the CSUQ for the PALT prototype were the highest number of users (N = 12) had a score ranging between 75 and 80. In all of the 19 items, except for item 9, most users agreed with the presented questions. More than half of the end users strongly agreed that the information and its organization was clear and easy to understand (items 11 = 42 (54.5%), 15 = 41(53.2%), 13 = 40 (51.9%) (Table 2). These findings align with reported means of the items (3.09 and 4.32). When categorizing the total CUSQ scale, 75% of users were very satisfied. The overall Cronbach’s alpha was 0.957 which indicates a high level of internal consistency, and hence reliability (Table 3).

Regarding construct validity, EFA yielded an orthogonal varimax rotation for two factor structure (p < 0.0001). Factor 1 (F1) received 73.7% (14 items) of item loadings with threshold of 0.60 in bold, whereas only 4 items loaded on Factor 2 (F2). Question (10) did not meet the threshold (Table 4). A mean of 75.68 ± 14.1 8 out of 95 suggests a high satisfaction level (Table 5). Table 6, summarizes a comparison of factors loading per items between the current study and Lewis’ standardization data. Ten items (1 – 10) (53%) were comparable to Lewis CSUQ. Items (11 – 15) loaded on factor 1 in the current study contrary to loading on 2 in Lewis’.

![Fig. 2. School testing setting using the Paediatric Arabic Language Test app prototype.](image)

| Table 1 | End-users by level of experience and age range. |
|---------|-----------------------------------------------|
| Level of experience | No | Age range |
| Junior student clinicians | 57 | 19–20 |
| Senior student clinician (Level 8) | 18 | 21–22 |
| Clinicians (+7 years' experience) | 2 | 34–37 |
| Total | 77 | |

4486
Items (16 – 18) loaded on factor 2 compared to 3 in Lewis’. Item 19 though loaded on factor 1 in this study, was not available comparison in Lewis’ study.

Possible variables may have impacted the direction of item-factor association. These may include use of a 5-point Likert scale, cross-cultural and clinical assessment field testing. To statistically reduce the possible effect of the 5-point Likert scale, three conversion methods were applied. The z-standardization score is a procedure used to transform the absolute rating values (1 = Strongly Disagree to 5 = Strongly Agree) to relative scores that reflect the response’s rank related to the rank of all responses. The second method was the percent of maximum possible (POMP) (Fischer and Milfont, 2010). Advantages of POMP is maintaining the multivariate and covariance metric of the rating scores. The converted scores on the 5-point scale to a metric value of 0 (minimal possible) to 100 (maximum possible) represent the possible maximum score (Moeller, 2015). The third method was IBM transformation method to convert a 5-point Likert scale to a 7-point Likert scale (IBM, 2020).

PCA with Varimax rotation was run on the questionnaire converted scores. Across all methods, items loaded on two factors with the same values as in Table 4. Although the first factor included the learnability items suggested by previous work (Questions 1–8), it also contained questions that do not reflect learnability (Questions 9, 11–15) (Table 7).

4. Discussion

Rating usability at the prototype phase of mHealth apps seemed useful and cost-effective as it informs healthcare practitioners about the apps feasibility and improvement prior to commercial stage (Lewis, 1995). Further, application of cross-cultural usability scales proved to yield different results from the original validation data. Although data in the current study showed overall high satisfaction, the scale yielded different constructs and was unidimensional. Finally, there was no difference among scale results when three conversion methods were applied confirming the different results.

The high rating was expected given that the design was carefully constructed using evidenced based theory and clinical practice. The feedback on the usability of the prototype for clinical assessment of speech pathology indicated that end-users agreed on the effectiveness, efficiency, and were satisfied with the prototype for the intended purpose of diagnostic assessment.

An implication of the findings is that usability scales may depend on cross-cultural population (Lewis, 2014; Marcus, 2007). The divide of the items might indicate that the app prototype is evaluated by end users along the construct of “usability benefit” (easy-to-use, highly integrated in the clinical assessment of the target area, well interactive, transparent) and “usability limitations” (automated error recover).

A further factor to consider is the context of field research in which the scale was applied. In this study, the usability rating was tested for a diagnostic mHealth app prototype in a non-clinical setting. Results supported that the design had a user interface that allows effective operation and information processing that aids the user on conducting clinical assessment in non-laboratory field settings. The clinicians were able to learn to use the system easily with minimal required training. The study provides evidence of commercial potential and value of the app prototype for clinical assessment or outcomes. The development of diagnostic apps may help to relieve financial and access difficulties by leading evaluation not only in local schools but also at home, utilizing commercially accessible devices with internet access.

Table 2
Percentages, means and SD of the CSUQ ratings for the PALT prototype (N = 77).

| Items | Strongly Disagree | Strongly Agree |
|---|---|---|
| 1 | 1(1.3%) | 4(5.2%) | 22(28.6%) |
| 2 | 0(0%) | 7(9.1%) | 13(16.9%) |
| 3 | 12(15.6%) | 15(19.5%) |
| 4 | 3(3.9%) | 2(2.6%) | 11(14.3%) | 31(40.3%) | 23(29.9%) | 27(35.1%) | 3.95 | 0.965 |
| 5 | 0(0%) | 5(6.5%) | 24(31.2%) |
| 6 | 3(3.9%) | 2(2.6%) | 19(24.7%) |
| 7 | 1(1.3%) | 4(5.2%) | 21(27.3%) |
| 8 | 16(20.8%) | 26(33.8%) |
| 9 | 8(10.4%) | 26(33.8%) |
| 10 | 3(3.9%) | 13(16.9%) |
| 11 | 6(7.8%) | 11(14.3%) |
| 12 | 11(13.3%) | 15(19.5%) |
| 13 | 6(7.8%) | 15(19.5%) |
| 14 | 5(6.5%) | 21(27.3%) |
| 15 | 10(13%) | 26(33.8%) |
| 16 | 6(7.8%) | 23(29.9%) |
| 17 | 10(13%) | 26(33.8%) |
| 18 | 8(10.4%) | 20(26%) |
| 19 | 5(6.5%) | 21(27.3%) |
| 20 | 2(2.6%) | 6(7.8%) | 23(29.9%) |
| 21 | 4(5.2%) | 14(18.2%) |
| 22 | 2(2.6%) | 6(7.8%) | 23(29.9%) |
| 23 | 10(13%) | 26(33.8%) |
| 24 | 8(10.4%) | 20(26%) |
| 25 | 5(6.5%) | 21(27.3%) |

Fig. 3. Distribution of the overall CSUQ scores for PALT prototype.
As for cultural obstacles, such as the theory that mHealth usability studies is limited to research contexts, were approached in this study by pointing out that the application of mHealth usability studies in a real-world framework, outside a specialist research department, is feasible (Sutherland et al., 2017; May and Erickson, 2014).

One of the limitations of this study was related to the measurement scale. A 5-point Likert scale was used instead of a 7-point Likert scale. However, statistics showed identical results. It is hypothesized that the high usability is due to the theoretical and clinical experience the developers engaged in constructing the PALT design relying on evidence-based research and practice specific to Saudi Arabic population and culture. Digital integration into paediatric clinical language assessment could support making speech pathology specialist assessments more accessible, informative and time efficient. Future studies should also implement other validated instruments to compare more usability dimensions.

5. Conclusions

The positively high usability rating of the prototype should inspire practitioners to learn how their designs may be evaluated at an early phase for cost effectiveness and improved quality. Of particular note to these clinical practitioners is the finding that good usability could be achieved as early as the prototype. Results of this study supported a two-factor representation of the CSUQ instead of a three-factor structure. Such findings will provide indications for the importance of cultural and context adaptation of usability scales. We hope these results may aid in the creation of a “gold standard” for clinical assessment tools in speech pathology as there are no pre-existing tools.

6. Data sharing statement

The data that supported the findings of this study will be available from the corresponding author upon reasonable request as a Microsoft Excel© datasheet.

7. Ethics approval and informed Consent

The study was approved by the committee of scientific research ethics at King Saud University (CAMS 129–36/37), and the Ministry of Education, Planning Department (No. 37972368). Consent to participate was obtained from the end users by the online questionnaire after explaining the objective of the study and confidentiality.

| Table 3 |
| --- |
| Categorizing percentages of the total CSUQ scale scores. |

| Lewis Usability scale categories | n (%) |
| --- | --- |
| Not satisfied (19–31) | 0 N 0% percent |
| Neutral (32–62) | 19 25% |
| Very satisfied (63–95) | 57 75% |

| Table 4 |
| --- |
| Varimax-rotated factor pattern for the principal factor analysis of the CSUQ for PALT. |

| Items | F1 | F2 |
| --- | --- | --- |
| 1 | 0.79 | 0.26 |
| 2 | 0.81 | 0.15 |
| 3 | 0.89 | 0.17 |
| 4 | 0.70 | 0.51 |
| 5 | 0.79 | 0.36 |
| 6 | 0.65 | 0.39 |
| 7 | 0.72 | 0.14 |
| 8 | 0.62 | 0.47 |
| 9 | 0.01 | 0.84 |
| 10 | 0.42 | 0.56 |
| 11 | 0.74 | 0.18 |
| 12 | 0.71 | 0.41 |
| 13 | 0.64 | 0.44 |
| 14 | 0.84 | 0.28 |
| 15 | 0.61 | 0.51 |
| 16 | 0.27 | 0.83 |
| 17 | 0.25 | 0.84 |
| 18 | 0.41 | 0.69 |
| 19 | 0.66 | 0.58 |

| Table 5 |
| --- |
| Average factor loading matrix for the CSUQ (N = 77). |

| Mean ± SD | Min-Max |
| --- | --- |
| Total Lewis Usability scale | 75.68 ± 14.18 | 47–95 |
| F1* | 57.36 ± 10.76 | 34–70 |
| F2* | 18.31 ± 4.43 | 8–25 |

*p-value < 0.0001

As for cultural obstacles, such as the theory that mHealth usability studies is limited to research contexts, were approached in this study by pointing out that the application of mHealth usability studies in a real-world framework, outside a specialist research department, is feasible (Sutherland et al., 2017; May and Erickson, 2014).

| Table 6 |
| --- |
| A comparison of factors loading of the CSUQ items between the current study and Lewis. |

| Item | Factors | The current study | Factors | Lewis (1995) |
| --- | --- | --- | --- | --- |
| 1 | 1 | 0.79 | 1 | 0.74 |
| 2 | 1 | 0.81 | 1 | 0.69 |
| 3 | 1 | 0.89 | 1 | 0.72 |
| 4 | 1 | 0.70 | 1 | 0.74 |
| 5 | 1 | 0.79 | 1 | 0.77 |
| 6 | 1 | 0.65 | 1 | 0.72 |
| 7 | 1 | 0.72 | 1 | 0.63 |
| 8 | 1 | 0.62 | 1 | 0.66 |
| 9 | 2 | 0.84 | 2 | 0.72 |
| 10 | 2 | 0.56 | 2 | 0.67 |
| 11 | 1 | 0.74 | 2 | 0.81 |
| 12 | 1 | 0.71 | 2 | 0.77 |
| 13 | 1 | 0.64 | 2 | 0.76 |
| 14 | 1 | 0.84 | 2 | 0.73 |
| 15 | 1 | 0.61 | 2 | 0.57 |
| 16 | 2 | 0.83 | 3 | 0.81 |
| 17 | 2 | 0.84 | 3 | 0.81 |
| 18 | 2 | 0.69 | 3 | 0.56 |
| 19 | 1 | 0.66 | N/A | 0.50 |
PCA with Varimax rotation of the CSUQ ratings for this study using z-standardization, POMP and IBM SPSS methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The author would like to thank all study participants.

Funding

This research project was supported by a grant from the “Research Centre of the Female Scientific and Medical Colleges,” Deanship of Scientific Research, King Saud University.

References

Buttle, F., 1996. SERVQUAL: review, critique, research agenda. Eur. J. Mark. 30, 8–32. https://doi.org/10.1108/03090569610105762.

Claris FileMaker Pro https://www.claris.com/filemaker/pro/. 2020.

Furlong, L., Morris, M., Serry, T., Erickson, S., 2018. Mobile apps for treatment of speech disorders in children: an evidence-based analysis of quality and efficacy. Front. Psychol. 9, 1389. https://doi.org/10.3389/fpsyg.2018.01389.

Krosnick, J., Presser, S., 2010. Question and Questionnaire Design In: Wright PMA, ed. Handbook of Survey Research 2nd Edition: Bingley: Emerald Group Publishing.

Lewis, J., 2014. Usability: Lessons Learned and Yet to Be Learned. International Journal of Human-Computer Interaction. 30, 663–684. https://doi.org/10.1080/10447318.2014.930311.

IBM, 2020. Transforming different Likert scales to a common scale. 2020; https://www.ibm.com/support/pages/node/422073.

Lewis, J.R., 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int. J. Human-Computer Interact. 7 (1), 57–78. https://doi.org/10.1080/10447319509526110.

Marcus A. (2005).2 chapter - User Interface Design’s Return on Investment: ed. Handbook of Survey Research 2nd Edition: Bingley: Emerald Group Publishing.

McKean C, and Bloch S. (2019). The application of technology in speech and language therapy. Int J Lang Commun Disord. 03; 54(2):157-158. PubMed ID 30811757

Moeller J. A word on standardization in longitudinal studies: don’t. Front Psychol. 2015;6:1389–1389. doi.org/10.3389/fpsyg.2015.01389.

Sutherland, R., Trembath, D., Hodge, A., et al., 2017. Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. Journal of Telemedicine and Telecare. 23 (1), 106–115. https://doi.org/10.1177/1357633X16655939.

Taylor, O.D., Armfield, N.R., Dodrill, P., Smith, A.C., 2014. A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. Journal of Telemedicine and Telecare. 20 (7), 405–412. https://doi.org/10.1177/1357633X14523388.