



## OPEN ACCESS

### EDITED BY

Andrea Greco,  
University of Bergamo, Italy

### REVIEWED BY

Jasmine Begeske,  
Purdue University, United States  
Xuan Tang,  
Guangzhou Huashang College, China

### \*CORRESPONDENCE

Matilde Spinoso  
✉ matilde.spinoso2@unibo.it

RECEIVED 09 September 2025

REVISED 19 February 2026

ACCEPTED 23 March 2026

PUBLISHED 07 April 2026

### CITATION

Spinoso M, Benassi M, Mazzoni N, Orsoni M, Stefanutti L, Anselmi P, de Chiusole D, Bacherini A, Pierluigi I, Garofalo S, Balboni G and Giovagnoli S (2026) Psychometric properties of the Usability and Acceptability Scale (UAS) for evaluating digital tools in children and adolescent users.

*Front. Psychol.* 17:1702085.

doi: 10.3389/fpsyg.2026.1702085

### COPYRIGHT

© 2026 Spinoso, Benassi, Mazzoni, Orsoni, Stefanutti, Anselmi, de Chiusole, Bacherini, Pierluigi, Garofalo, Balboni and Giovagnoli. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Psychometric properties of the Usability and Acceptability Scale (UAS) for evaluating digital tools in children and adolescent users

Matilde Spinoso<sup>1\*</sup>, Mariagrazia Benassi<sup>1</sup>, Noemi Mazzoni<sup>1</sup>, Matteo Orsoni<sup>1</sup>, Luca Stefanutti<sup>2</sup>, Pasquale Anselmi<sup>2</sup>, Debora de Chiusole<sup>2</sup>, Alice Bacherini<sup>3</sup>, Irene Pierluigi<sup>3</sup>, Sara Garofalo<sup>1</sup>, Giulia Balboni<sup>1</sup> and Sara Giovagnoli<sup>1</sup>

<sup>1</sup>Department of Psychology "Renzo Canestrari", Cesena Campus University of Bologna, Cesena, Italy,

<sup>2</sup>Department of Philosophy, Sociology, Education, and Applied Psychology, University of Padua,

Padua, Italy, <sup>3</sup>Department of Philosophy, Social Sciences, and Education, University of Perugia, Perugia, Italy

The rapid integration of digital tools in educational and clinical settings highlights the need for assessing their usability and acceptability, particularly among populations of developmental age. This study aims to validate the Usability and Acceptability Scale (UAS), a new scale tailored for children aged 4 to 18, derived from integration and an adaptation of the Technology Acceptance Model Scale and System Usability Scale. The UAS was administered to a sample of 908 participants. Results of Exploratory and Confirmatory Factor Analyses consistently supported a bifactorial model encompassing the two dimensions of usability and acceptability. Reliability was also assessed using Cronbach's  $\alpha$  and McDonald's  $\omega$ , with overall results indicating acceptable internal consistency. These findings suggested that the UAS is a valid and reliable questionnaire for evaluating digital tools among younger users, offering valuable insights for developers and educators aiming to create child-friendly technologies.

### KEYWORDS

acceptability, children and adolescents, digital tools, psychometric properties, usability, usability questionnaire, users scale

## Introduction

Nowadays, technological advancements are progressing at an unprecedented rate compared to previous times. Technology plays an increasing role in children's lives, shaping their learning, entertainment, and social interactions. In this context, the usability and acceptability of new technologies are particularly relevant for the effectiveness and implementation of new digital tools, such as computerized neuropsychological tests (Davis, 1989; Venkatesh and Davis, 2000). Concerning the latter, it is essential to ensure that technology not only enhances assessment quality but is also perceived as accessible and easily usable by children (Druin, 2002). Acceptability refers to the extent to which a user finds a computer system suitable, agreeable, and satisfactory. It is closely linked to the concept of usability, which measures the extent to which specified users can use a product to achieve specific goals with effectiveness, efficiency, and satisfaction in a defined context of use (ISO, 1998).

Although usability and acceptability are closely related constructs, they are conceptually distinct. Usability refers to the system's functional capacity to support effective and efficient task performance (performance-oriented; ISO, 1998). In contrast, acceptability reflects users' cognitive and motivational beliefs influencing their intention to engage with the system (motivation-oriented; Venkatesh and Bala, 2008). While both constructs involve the notion of "ease," usability concerns the objective support provided during task execution, whereas perceived ease of use within the acceptance framework reflects the user's subjective belief that interacting with the system requires little effort.

Assessing the acceptability and usability of digital technologies in children is crucial for several reasons. Children have distinct cognitive and motor skills compared to adults, which significantly influence their interaction with technology (Piaget, 1970; Diamond, 2000). Therefore, tools designed for children must be tailored to their specific needs and capabilities (Markopoulos et al., 2008; Hourcade, 2007). Additionally, children's attitudes toward technology can affect their willingness to engage with digital tools, impacting the effectiveness of educational and therapeutic interventions (Druin, 2002). Moreover, understanding the usability and acceptability of these technologies can help developers create more engaging and effective products, ultimately enhancing the overall user experience for children. To date, to effectively assess usability and acceptability various methods are employed, including usability testing methods (e.g., the System Usability Scale (SUS), and task completion rates; Brooke, 1996; Rubin and Chisnell, 2008), direct observation techniques (e.g., think-aloud protocols; Nielsen, 1994; van den Haak et al., 2003), heuristic evaluation based on established design principles (e.g., Nielsen's heuristics; Nielsen and Molich, 1990; Jeffries and Desurvire, 1992), cognitive walkthrough approaches (e.g., Polson et al., 1992; Wharton et al., 1994) and focus groups (e.g., contextual interviews and group discussions; Kuniavsky, 2003). Among these approaches, questionnaires represent one of the most widely used tools in the field (Brooke, 1996; Bangor et al., 2008). When applied to pediatric populations, these traditional methods require substantial adaptations to accommodate children's developmental characteristics. Specifically, the assessment of usability and acceptability in children draws on four main categories of methods: observational, verbalization, survey, and longitudinal approaches (Markopoulos et al., 2008). Observational methods range from passive to participatory, with the moderator choosing whether or not to engage with the participant and may be complemented by additional tools such as eye-tracking to monitor interface interaction (Masood and Thigambaram, 2015). Verbalization methods capture what participants say, either spontaneously or upon request, and include concurrent and retrospective think-aloud protocols (Donker and Markopoulos, 2002), active and robotic intervention, and post-task interviews, which are particularly useful for collecting immediate data (Bauw and Markopoulos, 2004).

Collaborative techniques such as co-discovery and peer tutoring are employed to facilitate more natural communication between children (Downey, 2007; Markopoulos and Bekker, 2003; Ognjanovic and Ralls, 2013). For interactive technology products, the Wizard-of-Oz method allows a moderator to control interactivity remotely and is particularly useful in early development stages (Höysniemi et al., 2004). Longitudinal methods such as diary studies allow observation of product use in naturalistic contexts over time, although their implementation in applied settings may be demanding (Markopoulos et al., 2008; Jacoby, 2011).

Survey methods include questionnaires and interviews, which can be combined to clarify ambiguous responses (Markopoulos et al., 2008). Child-friendly adaptations, such as the Fun Toolkit, increase engagement and reduce satisficing (Read, 2008; Read and MacFarlane, 2006; Vanette and Krosnick, 2014). Moreover, several adult usability questionnaires have been adapted for pediatric populations through visual supports and age-appropriate wording (Brooke, 1996; Bangor et al., 2008; Sim et al., 2006; Read, 2008; Barendregt et al., 2006; Kantosalo and Riihiahho, 2019; Banker and Lauff, 2022). Importantly, many of the available instruments rely on self-report methodologies. Although self-report measures are indispensable for capturing children's subjective experiences, their application in developmental populations requires careful methodological consideration. Children are more vulnerable than adults to response biases such as acquiescence, suggestibility, and social desirability (Bell, 2007; Soto et al., 2008). Developmental differences in cognitive, linguistic, and executive functioning may affect children's comprehension of item wording, response scales, and abstract constructs, potentially influencing response validity (Emerson et al., 2013). These challenges may be particularly salient in younger participants and in children with borderline or lower cognitive functioning, where reduced metacognitive monitoring and executive control can increase susceptibility to systematic response patterns (Nicolaidis et al., 2020). Difficulties with understanding may lead to incorrect or incomplete responses, while the introduction of support from another person when completing self-report measures may introduce certain types of bias, such as socially desirable responding that often lead children to provide answers aligned with perceived expectations (van de Mortel, 2008), as a consequence of respondent-assistant dynamics (Finlay and Antaki, 2012; Kramer et al., 2010).

Despite the wide range of available methods to assess usability and acceptability, a significant gap remains in the available instruments: currently, there are no brief, validated and standardized tool that integrates both dimensions and are specifically designed for children and adolescents aged 4 to 18 years. Existing instruments are either too complex, requiring metacognitive and linguistic skills that are not yet fully developed, or overly simplified, resulting in reduced discriminative power and evaluative comprehensiveness (Hourcade, 2007; Markopoulos et al., 2008). This limitation is particularly pronounced in the Italian context, where, to our knowledge, no questionnaires specifically developed for the developmental age population currently encompass both dimensions.

To address this need, the present study aims to create and validate a new questionnaire, the Usability and Acceptability Scale (UAS), specifically tailored for children and adolescents aged 4 to 18. By focusing on this age group, the UAS seeks to enhance our understanding of how children engage with technology, ultimately contributing to the development of more effective and engaging digital tools. The UAS is an adapted version of the Technology Acceptance Model (TAM; Venkatesh and Bala, 2008) and System Usability Scale (SUS; Brooke, 1996) questionnaires. In this study, we administered the UAS to assess the usability and acceptability of new digital tools for neuropsychological evaluation, such as MatriKS, a computerized test measuring fluid intelligence (de Chiusole et al., 2024).

Introduced by Davis (1989), TAM is one of the most influential models for understanding user acceptance of technology. TAM has been extensively applied in various contexts, including educational technology (Dizon, 2016; Lee et al., 2009; Ratna and Mehra, 2015; Teo et al., 2008), healthcare (Nikolaus et al., 2014), and consumer

software (Lilley et al., 2004), due to its easiness and predictive power (Davis et al., 1989; Lee et al., 2003). However, as technology evolves and new factors that influence user acceptance emerge, the need for more comprehensive models becomes evident. TAM2 (Venkatesh and Davis, 2000) extended the original model by including additional variables such as social influence and cognitive instrumental processes, further refining the understanding of technology adoption. Building upon this, TAM3 (Venkatesh and Bala, 2008) introduced an even more nuanced framework by integrating the determinants of perceived ease of use from the original TAM with the determinants of perceived usefulness from TAM2. TAM3 considers a wider range of factors, including computer self-efficacy, perceptions of external control, and perceived enjoyment, making it particularly suitable for evaluating complex and interactive technologies. This model provides a robust framework for assessing the likelihood of technology adoption by addressing both user perceptions and contextual factors that may influence technology acceptance (Davis et al., 1989).

The SUS (Brooke, 1996) is a widely used standardized questionnaire for measuring usability, specifically designed to generate a comprehensive single measure of perceived usability. It is known as a “quick” survey scale that allows practitioners to assess the usability of novel technologies easily and efficiently. Since its introduction, SUS has been incorporated into a variety of commercial usability toolkits and technology-based systems, including consumer software, websites, applications, and hardware products (Bangor et al., 2008; Blažica and Lewis, 2015; Brooke, 2013; Lewis, 2018; Marzuki et al., 2018).

The SUS items have been developed according to the three usability criteria defined by ISO 9241-11: (1) the ability of users to complete tasks using the system and the quality of the output of those tasks (i.e., effectiveness); (2) the level of resource consumed in performing tasks (i.e., efficiency); and (3) the users’ subjective reactions using the system (i.e., satisfaction).

The original SUS (Brooke, 1996) includes a mix of positive and negative items, and it was structured to control for acquiescence bias and to identify respondents who did not pay attention to the statements. However, several studies have highlighted the potential issues associated with the inclusion of both positive and negative items (Barnette, 2000; Stewart and Frye, 2004). For example, it can result in a reduction in internal reliability (Stewart and Frye, 2004), a distortion of the factor structure (Pilotte and Gable, 1990; Schriesheim and Hill, 1981), and an increase in the number of interpretation problems encountered when using the scale in cross-cultural contexts (Wong et al., 2003).

Moreover, mixed items may lead to difficulty in switching users’ response behaviours, and they can increase the cognitive load (Kortum et al., 2021). This could be particularly true for clinical and younger populations, whose executive functions may be impaired or still in development.

Notably, some authors suggested that questionnaires for usability assessment should avoid the inclusion of a mixture of positive and negative items and that researchers who do not specifically need to use the standard SUS should consider using the positive version to reduce the likelihood of response or scoring errors (e.g., Lewis, 2018). A version of the SUS that includes only positively worded items has been created (Sauro and Lewis, 2011) and it has been found to have satisfactory reliability, validity, and sensitivity (Lewis, 2018). Thus, the positive SUS emerges as a valuable alternative to the standard SUS, offering the advantages of reduced cognitive load and shifting ability.

Since 2014, several translations and psychometric evaluations of SUS have been published, including Arabic (AlGhannam et al., 2017), Danish (Hvidt et al., 2020), Chinese (Wang et al., 2020), French (Gronier and Baudet, 2021), Italian (Borsci et al., 2009), Malay (Marzuki et al., 2018), Persian (Dianat et al., 2014), Polish (Borkowska and Jach, 2017), Portuguese (Martins et al., 2015), Slovene (Blažica and Lewis, 2015), and other languages. In addition, translations of SUS into German (Perrig and von Felten, 2025), and Finnish (Raita and Oulasvirta, 2011) have been conducted, although there is no available data regarding their psychometric properties.

Despite its wide adoption a significant need for validations in different populations remains (Lewis, 2018).

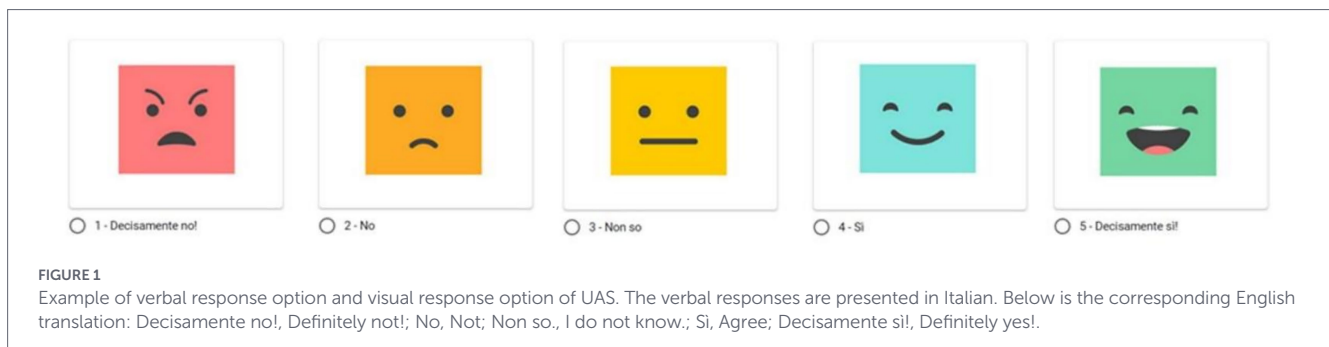
Based on these premises, we developed and validated the UAS, a new and comprehensive Italian scale designed to assess both the usability and acceptability of a digital tool in the developmental population. Starting from assessing face validity and content validity of the scale, the present investigation aimed to validate the UAS in a group of children aged 4–18. We investigated the construct validity of the UAS, examining the factorial structure of the scale using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Then we investigated the reliability of UAS measurement, checking the internal consistency of Cronbach’s  $\alpha$ , McDonald’s  $\omega$ , and corrected item-total correlation. Moreover, we verified if the factorial structure was invariant for sex and age. This study would provide evidence about the reliability and validity of this new tool for assessing the usability and acceptability of digital technologies in children and adolescents, thereby enhancing the development and implementation of child-friendly digital assessments.

## Materials and methods

### The Usability and Acceptability Scale (UAS)

The UAS is an eight-item questionnaire, based on the Italian adaptations of the SUS (Brooke, 1996) and TAM (Davis, 1989; Venkatesh and Bala, 2008). It has been specifically designed to assess usability and acceptability in children and adolescents. It has been structured to achieve brevity and ease of comprehension, simplifying the language and using emoticons to make it easier for the younger users to understand and answer the questions. In the UAS, participants are asked to rate their agreement with eight statements using a 5-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree.” Each statement is presented in a child-friendly format using simple language accompanied by icons. Specifically, the verbal response options were paired with coloured emoticon images, this visual aid was designed to increase comprehension and engagement in the rating process (see Figure 1).

In the present study, we asked 10 experts to evaluate the content validity of the questionnaire. Experts were chosen based on their expertise in the field of usability and acceptability and their experience with children in neuropsychological assessment. To ensure a structured and replicable evaluation process, each expert received a standardized package via email, including: a background questionnaire, the preliminary version of the UAS scale, an introductory letter describing the aims of the study and the target population, and a structured content validation form. Experts were asked to independently evaluate each item with respect to multiple content-related criteria, including relevance to the construct, clarity of wording, appropriateness for the target population, and comprehensibility for children. In addition to quantitative ratings,



experts were invited to provide qualitative comments and to suggest item modifications, deletions, or the addition of new content areas.

All completed forms were returned electronically and systematically reviewed by the research team. Expert feedback was analyzed both quantitatively (i.e., inspection of ratings across experts) and qualitatively (i.e., thematic analysis of open-ended comments). The item revision process followed a set of predefined decision rules aimed at maximizing content validity: items were retained when they were consistently judged as highly relevant and comprehensible; items were revised when experts identified issues related to wording, developmental appropriateness, or ambiguity; and items were removed when they were judged as redundant, insufficiently relevant to the target constructs, or inappropriate for children. When multiple experts suggested similar changes, these recommendations were prioritized in the revision process.

Following this expert-based review, the questionnaire underwent a structured item reduction and refinement process. The original questionnaire consisted of  $N = 22$  items derived from the System Usability Scale (SUS) and the Technology Acceptance Model (TAM3) (see [Supplementary Tables S1 and S2](#) for the original items). Based on expert evaluations, the scale was reduced to a more parsimonious version while preserving coverage of the core dimensions of usability and acceptability. This process resulted in the selection, modification, and refinement of items judged to best represent the intended constructs and to be most suitable for the developmental level of the target population. The final version of the UAS consists of eight items: four items assessing the usability dimension adapted from the SUS, and four items evaluating the acceptability dimension, specifically, the perceived ease of use as a component of acceptability, derived from TAM3 items.

In line with the recommendations of [Brooke \(1996\)](#), who advocated for a flexible application of the SUS items according to the specific context of interest, only four out of 10 SUS items were considered. Specifically, consistent with the positive SUS version ([Sauro and Lewis, 2011](#); see [Supplementary Table S1](#), for the complete version), only positive worded items were included. Namely, Items 1 (“I think that I would like to use the...frequently”), 2 (“I found the...to be simple”), 3 (“I thought the...was easy to use”), and 7 (“I would imagine that most people would learn to use the...very quickly”), as experts judged these items to be more easily comprehensible and cognitively less demanding for the targeted audience of children, thereby increasing developmental appropriateness and reducing potential response bias associated with negatively worded items. These source items were subsequently adapted and reworded for the specific context and target population, resulting in the following final usability items: 1 “I would gladly take the test again,” 2 “The test was very easy,” 3 “I think that anyone could take this test,” and 4 “It was easy to understand what to do during this test” (as shown in [Table 1](#)).

A similar theoretically and developmentally informed approach was applied to the selection of acceptability items. The original TAM version consists of 12 items, with six evaluating perceived usefulness (PU) and six assessing perceived ease of use (PEU; see [Supplementary Table S2](#), for the complete list of original items). PU refers to the degree to which a person believes that technology will enhance job performance, while PEU is defined as the extent to which a person believes that using technology will be effortless ([Davis, 1989](#)). Given that the UAS was designed to assess the usability and acceptability of MatriKS, a new digital assessment tool for assessing fluid intelligence intended to support the job performance of clinicians and experimenters, rather than to improve the performance of the tested participants. PU items were judged by experts as conceptually inappropriate for the test-taker population. Accordingly, only PEU related items were considered relevant for inclusion.

Within the PEU items, a further expert-guided selection was performed to retain items that best captured children’s direct experience with the tool while minimizing redundancy and cognitive load. Four PEU items were ultimately selected: the UAS questionnaire selectively incorporated just four items. Specifically, Items 5 (“Using... would make it easier to do my job”) and 6 (“I would find...useful in my job”) pertained to the perceived ease of use for oneself and others, Item 9 (“My interaction with the...is clear and understandable”) and Item 7 (“I find the...to be easy to use”) captured core aspects of perceived ease of use, Item 11 (“I find using the...to be enjoyable”) focused on perceived pleasantness/fun, and Item 12 (“Assuming I had access to the..., I intend to use it”) addressed intentionality. These source items were subsequently adapted and reworded for the specific tablet-based testing context and for child respondents, resulting in the following final acceptability items: 5 “I think that taking the test on the tablet was easy,” 6 “I think that it would be easy for anyone to take the test on the tablet,” 7 “I felt good taking the test on the tablet,” and 8 “I would like to always use the tablet for this test.” The full mapping between source and final UAS items is reported in the [Supplementary Tables S1 and S2](#). Additionally, to the expert-based content validation, face validity was further examined through direct input from the target population. Face validity was defined as the extent to which the scale’s purpose is apparent to the target users and appears to be an appropriate measure for its intended purpose ([Salkind, 2010](#); [Mosier, 1947](#)), encompassing aspects such as feasibility, readability, consistency in style and formatting, and the clarity of the language used ([DeVon et al., 2007](#)). A sample of  $N = 10$  participants from the target population was asked to select the statement they thought best reflected the construct in question and to change the wording, delete statements, or add new aspects they thought relevant. Participants were also invited to suggest wording changes, delete unclear statements, and propose additional aspects they considered relevant. Feedback from this phase was used to further refine item wording and to ensure that the final version of the UAS was

TABLE 1 Items of the final version of Usability and Acceptability Scale (UAS).

UAS scale	Italian version	English version
Item 1	Rifarei volentieri la prova	I would gladly take the test again
Item 2	La prova è stata molto facile	The test was very easy
Item 3	Penso che tutti possano fare questa prova	I think that anyone could take this test
Item 4	È stato facile capire cosa fare durante la prova	It was easy to understand what to do during this test
Item 5	Penso che fare la prova sul tablet sia stato semplice	I think that taking the test on the tablet was easy
Item 6	Credo che per tutti sia facile fare la prova con il tablet	I think that it would be easy for anyone to take the test on the tablet
Item 7	Sono stato bene a fare le prove sul tablet	I felt good taking the test on the tablet
Item 8	Vorrei usare sempre il tablet per fare questa prova	I would like to always use the tablet for this test

not only theoretically sound but also developmentally appropriate and easily interpretable by children. Overall, the iterative combination of expert review and target-user feedback resulted in a concise, developmentally sensitive instrument with enhanced evidence of content and face validity, while maintaining conceptual coverage of the usability and acceptability constructs relevant to the evaluation of MatriKS. The adaptation process, including linguistic simplification and item reduction, was finalized prior to the factorial analyses. The EFA and CFA were conducted on a split sample using the same predefined version of the scale, and no post-hoc modifications were made following the EFA.

The final adapted version of the scale (UAS) is publicly available and is freely accessible for research purposes.

## Procedure

In this study, we administered the UAS to assess the usability and acceptability of MatriKS, a new digital tool for the assessment of fluid intelligence based on the theory of knowledge structures (Doignon and Falmagne, 1985; Falmagne and Doignon, 2010; Heller and Stefanutti, 2024). MatriKS is available in two versions, differentiated according to participants' age, namely between 4 and 11 years old and 12 or more years old. Participants completed a different version of MatriKS according to their chronological age. The test is part of PsycAssist,<sup>1</sup> a platform for the assessment of neuropsychological functioning (de Chiusole et al., 2024).

Participants completed first MatriKS and then the UAS using a tablet (iOS operating system with a 10.9-inch screen). All participants completed the questionnaire independently, except for the youngest children, aged 4 to 6. For these children, examiners assisted by reading the items aloud, as the children's reading skills were not yet fully

developed. Children then responded autonomously to the items, with visual support provided by emoticons.

## Participants

The UAS was administered to a total sample of  $N = 908$  participants of the general population aged 4 to 18 (Female = 53%, Male = 47%, Mean age:  $10.53 \pm 3.50$ ). The descriptive statistics of the general sample are reported in Table 2. Participants were randomly divided into two groups to conduct the EFA (EFA group,  $n = 359$ ) and CFA (CFA group,  $n = 549$ ). The two groups did not differ in terms of age ( $t_{(906)} = -0.29$ ;  $p = 0.79$ ), sex ( $\chi^2_{(1)} = 1.30$ ;  $p = 0.25$ ), or educational level ( $\chi^2_{(3)} = 0.63$ ;  $p = 0.89$ ). The sample size of the two groups was established *a priori* according to the minimum criteria to have a subject to an item ratio of 10:1 in the EFA (Nunnally, 1978) and at least 10 observations for each freely estimated model parameter in the CFA (Kline, 1998, 2023). The data were collected in different regions of the North, Centre, and South of Italy. Schools of all levels were involved, from kindergarten to high school. The school principals were contacted in advance to inform them about the aim of the project and to inquire about their willingness to participate in the project. Subsequently, the parents or legal guardians of the children enrolled in the participating classrooms were provided with the informed consent in accordance with the Declaration of Helsinki recommendations (Williams, 2008). Only children whose parents provided informed consent were included in the study, and subsequently, the children themselves decided whether or not to participate. The exclusion criteria were the presence of motor, visual, and auditory impairments that prevented the participants from completing the task and the questionnaire. An expert team of psychologists handled the assessments and managed communications with both teachers and parents.

The study was approved by the ethical committee for the psychological research of the University of Padua (protocol code E047A9B0520E9732A8365DC72335EE90/ date of approval: 8 July, 2022).

## Data analysis

All statistical analyses were performed on the variables related to the final version of the UAS (see Table 1 for details on items) using IBM SPSS 25 (SPSS Inc., Chicago, IL, United States) and JASP version 0.18.3.0 (JASP Team, 2024).

As suggested by Tabachnick and Fidell (2013), for the EFA and CFA groups we checked for the presence of univariate outliers (scores more than 3.29 standard deviations above or below the corresponding group mean were excluded), normalized the UAS total score distribution, and checked for the presence of multivariate outliers. Each time we excluded participants, we normalized the UAS score distribution.

## Exploratory Factor Analysis

The Exploratory Factor Analysis (EFA) was conducted using the Weighted Least Squares (WLS) technique, which aims to capture most of the variance across a set of variables with a reduced number of factors (Fabrigar and Wegener, 2012). An oblimin rotation was applied, suitable when factors are expected to be correlated (Fabrigar et al., 1999; Costello and Osborne, 2005).

<sup>1</sup> <https://psycassist.fisppa.unipd.it/en/>

TABLE 2 Descriptive statistics of the total sample of participants ( $N = 908$ ) and Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) groups.

Variables	Total sample ( $n = 908$ )	EFA group ( $n = 359$ )	CFA group ( $n = 549$ )
<b>Sex (<math>n</math>)</b>			
M-F	47-53	49-51	46-54
<b>Age (years)</b>			
Range	4-18	4-18	4-18
M (SD)	10.52 (3.50)	10.01 (3.45)	10.91 (3.48)
<b>Educational level (<math>n</math>)</b>			
Kindergarten	107	42	65
Primary school	346	142	204
Middle School	251	98	153
High School	204	77	127

M-F, male-female; M (SD), mean (standard deviation).

Parallel analysis, scree test, incremental variance, and interpretability of the pattern of factor loadings were employed to choose the number of factors to retain (Fabrigar et al., 1999). Items with loadings greater than 0.40 and cross-loadings less than 0.10 were considered for inclusion in a factor.

## Confirmatory Factor Analysis

The goodness of fit of the factorial structure of the scale identified in the EFA was tested. Maximum likelihood estimation with a mean-adjusted Chi-square test (MLM estimator), which is robust to non-normal score distributions, was used. The metric of the latent variables was set by fixing the factor loading of the first item to one, for each factor. Overall model fit was determined by using the Satorra-Bentler scaled Chi-square statistic ( $S-B\chi^2$ ), robust comparative fit index (rCFI), robust root mean square error of approximation (rRMSEA) with associated 95% confidence intervals (CIs), and standardized root mean square residual (SRMR) (Schermelleh-Engel et al., 2003). Values close to 0.95 for rCFI, smaller than 0.05 for rRMSEA, and smaller than 0.08 for SRMR suggest a reasonable fit (Byrne, 2013).

The factorial structure found in EFA was compared with alternative nested models, which were theoretically plausible. To this aim,  $\Delta S-B\chi^2$  and  $\Delta rCFI$  were used as fit indices. To indicate that the null hypothesis of equivalence should be rejected (i.e., the EFA factorial structure model had a better fit than the alternative model), a significant  $\Delta S-B\chi^2$  and a value of  $\Delta rCFI$  (which is less affected by sample size) higher than 0.01 are required (Cheung and Rensvold, 2002). Finally, Akaike's information criterion (AIC) was calculated, with a lower value indicating a better fit of the model to the data (Schermelleh-Engel et al., 2003; Sterba and Pek, 2012).

## Measurement invariance

To verify the measurement invariance of the UAS, we tested the model across the sex of participants (female vs. male) and across three age groups (4-6 years, 7-11 years and 12-18 years), performing three levels of invariance: configural, metric, and scalar (i.e., same item intercepts) (Vandenberg and Lance, 2000). Measurement invariance was assessed specifically within the CFA subsample.

For the assessment of configural invariance, no constraints were imposed, enabling a test of whether the pattern of fixed and freely estimated parameters remained consistent across sub-groups. Regarding the metric invariance (weak invariance), the factor loadings of each item on the corresponding factor (i.e., the scale unit and the metric of each item) were constrained to be invariant across subgroups. Moreover, we performed the scalar invariance (strong invariance) where the factor loadings plus the intercepts of each item on the corresponding factor (i.e., the origin of the scale of each item) were constrained to be invariant (Byrne, 2008; Vandenberg and Lance, 2000).

## Sensitivity analyses

To examine potential effects of administration heterogeneity, we conducted sensitivity analyses comparing examiner-assisted administration (children aged 4-6 years) with non-assisted administration. Full details of these analyses are provided in the Supplementary materials.

## Internal consistency

As a measure of internal consistency Cronbach's  $\alpha$  (cutoff  $\geq 0.70$ ; Nunnally, 1978), McDonald's  $\omega$  (cutoff  $\geq 0.70$ ; McDonald, 1999), and corrected item-total correlations (cutoff  $\geq 0.30$ ; Norman and Streiner, 2008) were computed for each factor and for the total scale (Dunn et al., 2014; Sočan, 2000). These measures were assessed within the total sample.

## Results

### Exploratory Factor Analysis (EFA)

The EFA was conducted on the first subsample ( $n = 359$ ). Using parallel analysis, scree test, incremental variance, and interpretability of item factor loadings, all methods consistently extracted two inter-related factors, which were then rotated using oblimin rotation. The total explained variance was 42%. All items met the inclusion criteria for factor retention, with loadings greater than 0.40 and cross-loadings less than 0.10. Table 3 presents the item loadings for the two-factor solution (Model 1). The first factor consists of four items related to usability (items 1, 2, 3, and 4), while the second factor includes four items representing acceptability (items 5, 6, 7, and 8). Skewness and kurtosis computed on the two-factor scores indicated approximately normal univariate distributions, as all values were lower than  $|2|$  (Tabachnick and Fidell, 2013); skewness ranged from  $-1.347$  to  $0.149$  ( $SE = 0.18$ ), and kurtosis ranged from  $-0.853$  to  $3.153$  ( $SE = 0.357$ ).

### Confirmatory Factor Analysis (CFA)

The two-factor model selected in the EFA was tested on the second subsample ( $n = 549$ ) using CFA. Results indicated marginal to acceptable fit to the data, with all indices close to the expected value:  $S-B\chi^2 = 89.76$ ,  $p < 0.05$ ; rCFI = 0.92; rRMSEA = 0.082, 95% CI [0.066, 0.100]; SRMR = 0.044. It should be noted, however, that the rCFI value does not reach the conventional threshold of 0.95 and the rRMSEA slightly exceeds the recommended cutoff of 0.08, indicating a marginal—yet still acceptable—model fit. Overall, the pattern of indices suggests an adequate representation of the data. Aiming to evaluate possible alternative

TABLE 3 Exploratory Factor Analysis (EFA) results.

EFA factor loadings (n = 359)			
Item content	Mean ± SD	F1	F2
1. I would gladly take the test again	3.95 ± 0.88	<b>0.694</b>	0.010
2. The test was very easy	3.59 ± 0.98	<b>0.771</b>	-0.031
3. I think that anyone could take this test	3.71 ± 0.97	<b>0.456</b>	0.002
4. It was easy to understand what to do during this test	3.94 ± 0.96	<b>0.689</b>	0.055
5. I think that taking the test on the tablet was easy	4.34 ± 0.67	0.100	<b>0.584</b>
6. I think that it would be easy for anyone to take the test on the tablet	3.81 ± 0.98	-0.047	<b>0.548</b>
7. I felt good taking the test on the tablet	4.46 ± 0.61	0.082	<b>0.641</b>
8. I would like to always use the tablet for this test	4.15 ± 0.87	-0.091	<b>0.661</b>
<b>Eigenvalues after oblimin rotation</b>		1.8 (23%)	1.1(19%)

Bold values represent the factor loadings associated with the two factors.

models explaining specificity or overlapping between the investigated domains, the two-factor model obtained by EFA (Model 1, where the items of the questionnaire cluster around domains of usability and acceptability) was compared to a one-factor model solution (Model 2, representing unique general factor in which all items of the UAS were loaded on a single dimension). The CFA findings strongly support the validity of the UAS's two-factor structure showing that Model 1 was the most representative model of the questionnaire structure with the best fit and better results than the alternative nested model:  $\Delta S-B\chi^2$  ( $\Delta df$ ) range = 253.718 (20)–89.76 (19);  $\Delta rCFI$  range = 0.72–0.91. Concerning the other parsimony fit indices, it is possible to note as the two-factor model (Model 1) showed the lowest values for AIC (Model 1 = 10424.668; Model 2 = 10586.617) confirming that Model 1 turned out to be the model that best fits the data and best explains the dimensionality of the analyzed data. All factor loadings were statistically significant. Each item loaded highly (>0.50) and significantly ( $p < 0.001$ ) on its designated factor, with factor loadings from 0.50 to 0.70 (see Figure 2). Skewness and kurtosis computed on the CFA subsample indicated approximately normal univariate distributions, as all values were lower than |2| (Tabachnick and Fidell, 2013); skewness ranged from -1.153 to -0.338 (SE = 0.104), and kurtosis ranged from -0.735 to 1.128 (SE = 0.208).

## Measurement invariance across sex

The two-factor model exhibited full configural, metric, and scalar invariance (Table 4 presents the results of the sex invariance test, and fit indices for the model across males and females). All factor loadings were statistically significant and comparable between groups ( $p < 0.001$ ). The models assessing configural, metric, and scalar invariance demonstrated good fit indices that did not vary significantly from one model to the others. The UAS showed configural, metric, and scalar invariance across sexes (see Table 4), indicating that latent structure, factor loadings, and intercepts are equivalent for both groups.

## Measurement invariance across age

Three nested models with progressively stringent constraints were evaluated to test the invariance across three age groups 4–6 years (preschool/early childhood), 7–11 years (middle childhood), and 12–18 years (adolescence): configural invariance (Model 1), metric invariance (Model 2), and scalar invariance (Model 3). All factor loadings were statistically significant within each group and showed comparable magnitude across age groups. Table 5 presents the results of the age invariance analyses and model fit indices.

Results demonstrated acceptable configural invariance (GFI = 0.997, NFI = 0.830, SRMR = 0.055, RMSEA = 0.091), indicating that the basic two-factor structure is consistent across developmental stages. Metric invariance was also supported, as constraining factor loadings to equality did not significantly worsen model fit ( $\Delta\chi^2(15) = 21.95, p = 0.109$ ). This finding suggests that the relationship between observed indicators and latent constructs is stable across age groups and that the questionnaire items retain comparable meaning across developmental periods. In contrast, scalar invariance was not supported, as constraining item intercepts to equality resulted in a significant decrease in model fit ( $\Delta\chi^2(16) = 292.39, p < 0.001$ ). Fit indices for the scalar model (GFI = 0.991, NFI = 0.804, PNFI = 0.479, SRMR = 0.220, RMSEA = 0.152) further indicated reduced model fit, suggesting that item intercepts may differ across age groups.

## Internal consistency

The internal consistency of the UAS was assessed using both Cronbach's  $\alpha$  and McDonald's  $\omega$ , yielding promising results. The overall reliability was found to be acceptable, with Cronbach's  $\alpha$  at 0.74 and McDonald's  $\omega$  at 0.77, suggesting that the scale is internally consistent. When examining the individual factors, Factor 1 (F1) showed solid reliability with McDonald's  $\omega$  at 0.71 and Cronbach's  $\alpha$  at 0.70. However, Factor 2 (F2) had slightly lower internal consistency, with McDonald's  $\omega$  at 0.65 and Cronbach's  $\alpha$  at 0.66. While these values indicate somewhat lower reliability, they still fall within acceptable ranges, suggesting that both factors can be considered reliable, though further investigation might be warranted for Factor 2.

## Sensitivity analyses

Sensitivity analyses examining potential effects of administration heterogeneity (examiner-assisted for children aged 4–6 years vs. non-assisted administration) showed no significant differences in the total usability score. Small but statistically significant differences emerged at the factor level, in opposite directions across F1 and F2. Reliability indices were comparable across modalities, indicating

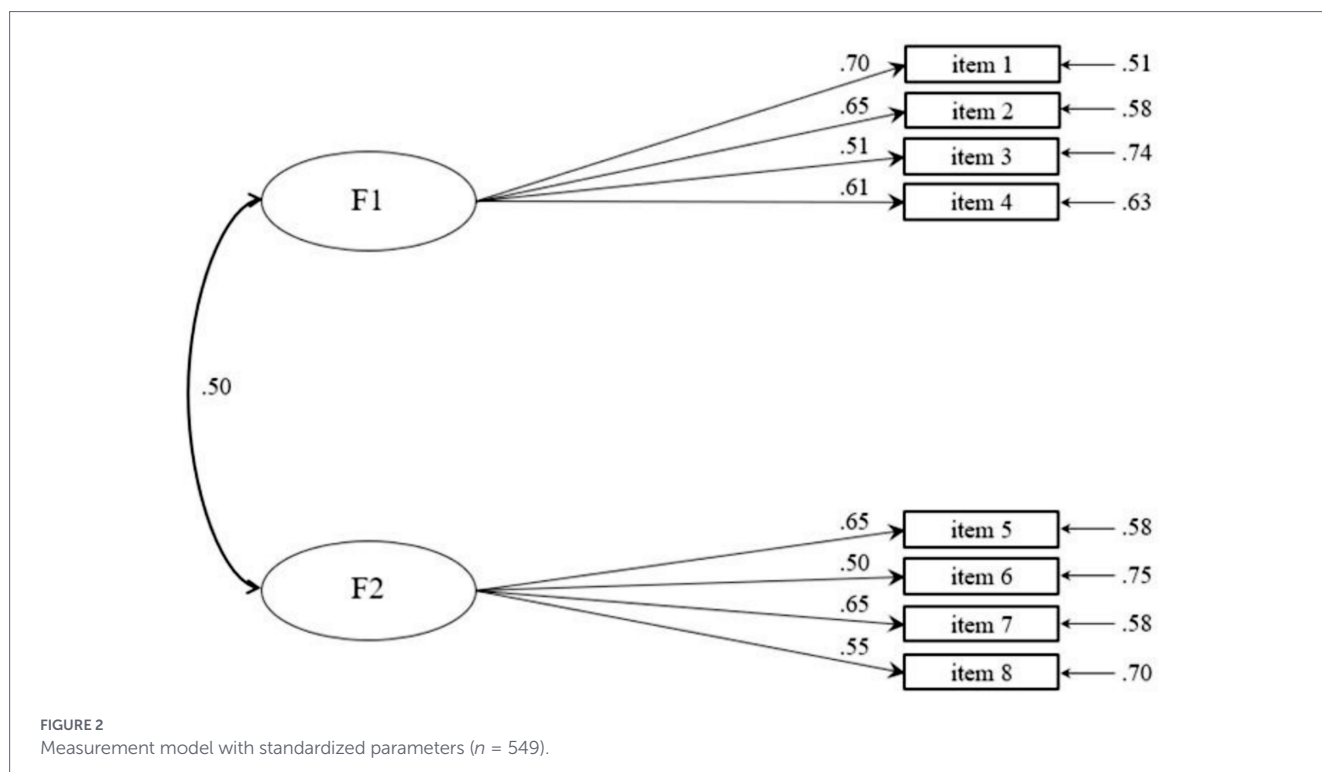


TABLE 4 Results of the measurement invariance test across groups (sex: male vs. female).

Model	GFI	NFI	PNFI	SRMR	RMSEA
Model (a)—configural: factor structure constrained to be equal	0.998	0.881	0.598	0.050	0.082
Model (b)—metric: factor loadings constrained to be equal	0.998	0.876	0.735	0.052	0.072
Model (c)—scalar: item intercepts constrained to be equal	0.997	0.860	0.844	0.050	0.069

GFI, Goodness of Fit Index; NFI, Normed Fit Index; PNFI, Parsimony-Adjusted Normed Fit Index; SRMR, standardized root means square residual; RMSEA, root mean square error of approximation.

TABLE 5 Results of the measurement invariance test across age groups (4–6, 7–11, 12+ years).

Model	GFI	NFI	PNFI	SRMR	RMSEA
Model (a)—configural: factor structure constrained to be equal	0.997	0.830	0.563	0.055	0.091
Model (b)—metric: factor loadings constrained to be equal	0.997	0.804	0.689	0.067	0.084
Model (c)—scalar: item intercepts constrained to be equal	0.991	0.804	0.479	0.220	0.152

GFI, Goodness of Fit Index; NFI, Normed Fit Index; PNFI, Parsimony-Adjusted Normed Fit Index; SRMR, Standardized root means square residual; RMSEA, root mean square error of approximation.

stable internal consistency. Complete statistical results and a more extensive discussion of these findings are provided in the [Supplementary materials](#).

## Discussion and conclusion

This study aimed to validate and analyze the psychometric properties of the Usability and Attitude Scale (UAS), a questionnaire

designed to assess the usability and acceptability of digital technologies among children and adolescents.

Our findings from EFA and CFA suggest that the UAS is a reliable and valid tool for evaluating usability and acceptability among children aged 4 to 18 years. The EFA revealed a clear factor structure for the UAS, indicating distinct dimensions for usability and acceptability. The results showed that the items loaded appropriately onto the hypothesized factors, with eigenvalues supporting a two-factor solution. This is consistent with the theoretical foundations of the Technology Acceptance Model (TAM) and System Usability Scale

(SUS), which posit that usability and attitude toward technology are critical determinants of user acceptance. The CFA confirmed the two-factor model structure identified in the EFA, with fit indices indicating a good model fit. These findings reinforce the scale's structural validity and suggest that the UAS is a reliable instrument for measuring the constructs it was designed to assess. The internal consistency, as indicated by Cronbach's  $\alpha$ , was overall satisfactory, providing further evidence of the reliability of the scale. However, with respect to the second factor, reliability indices were slightly below the conventional threshold and should therefore be interpreted with some caution. This may be partly attributable to the reduced number of items comprising this subscale, given that internal consistency coefficients such as Cronbach's  $\alpha$  are sensitive to scale length. Nevertheless, this does not substantially undermine the overall reliability of the scale, as the values remain close to recommended standards and the factor structure is theoretically coherent. With respect to the sensitivity analyses conducted to evaluate administration heterogeneity, findings indicate that administration modality does not meaningfully affect the overall usability score. Although small differences were observed at the factor level, these effects were limited in magnitude and did not impact internal consistency. Overall, the results support the robustness of the instrument's psychometric properties across administration formats, suggesting that the principal validation findings remain stable regardless of examiner assistance.

Our study aligns with previous research highlighting the importance of usability and acceptability in the adoption of digital technologies by children. The TAM framework has been extensively validated in adult populations, but its application to children is relatively novel. The successful adaptation and validation of the UAS demonstrate that these theoretical constructs are applicable and relevant to younger users as well.

The development and validation of the UAS for children represents a significant advancement in the assessment of digital technologies tailored for younger users. The UAS could be a valuable tool for developers, educators, and researchers engaged in the implementation of digital technologies for children. It provides a reliable measure of usability and acceptability, thus enabling the identification of strengths and improvement areas in digital products with the goal of better meeting the needs of young users. This can enhance the effectiveness of educational and therapeutic interventions delivered through digital platforms. For developers, the UAS provides insights that guide the design of user-friendly and engaging products. For educators and researchers, it assesses the impact of digital technologies on learning outcomes and engagement, promoting the evidence-based integration of technology into educational settings. Beyond structural validity, the present study examined the invariance across sex and age groups, a critical property for a scale intended for use across diverse developmental populations. Regarding sex invariance, the two-factor model showed full configural, metric, and scalar invariance across males and females, indicating that the latent structure, factor loadings, and item intercepts are equivalent between groups. These findings suggest that the UAS captures the same constructs in the same way regardless of sex, and that meaningful comparisons of latent mean scores between males and females are statistically justified. Regarding age invariance, results revealed a pattern of partial measurement invariance across the three developmental groups (4–6, 7–11, and 12–18 years). Configural invariance was supported, confirming that the two-factor structure of the UAS is consistent across developmental stages. Metric invariance was also established, as constraining factor loadings to equality across

groups did not significantly worsen model fit, indicating that the relationship between observed items and their underlying latent constructs is stable across developmental periods and that the questionnaire items retain comparable meaning from early childhood through adolescence. However, scalar invariance was not supported, as the equality constraints placed on item intercepts resulted in a significant deterioration of model fit. This finding suggests that children at different developmental stages may systematically endorse items at different levels, even when their standing on the underlying latent construct is equivalent. Such differences in item intercepts are theoretically plausible given the broad developmental span covered by the UAS and are consistent with changes in cognitive maturity, linguistic comprehension, and response tendencies across childhood and adolescence. Importantly, because metric invariance was achieved, the latent constructs retain the same substantive meaning across age groups, and the factorial structure of the scale remains stable. Caution is nonetheless warranted when making direct mean-level comparisons across developmental stages. Taken together, the invariance analyses provide robust evidence for the cross-group comparability of the UAS across sex and support its structural equivalence across developmental stages, strengthening its utility as a standardized assessment tool for children and adolescents.

Although the UAS demonstrates strong psychometric properties, we acknowledge some study limitations that warrant consideration. The data collection, though in diverse regions, was limited to Italy, and cultural factors may influence the generalizability of the findings. Therefore, future research should explore the applicability of the UAS in different cultural contexts to ensure its broader validity. A further limitation concerns the specific digital technology used for validation. The UAS has been tested exclusively with MatriKS, and additional studies are needed to verify its applicability and psychometric properties with other digital tools. Testing the scale across diverse digital platforms and applications will be essential to establish its generalizability and confirm its utility as a reliable measure of usability and acceptability for children's digital technologies.

Moreover, convergent and criterion validity were not assessed in this study. Currently, in the Italian context, no brief and easily administrable instruments exist for developmental populations that simultaneously capture both usability and acceptability of digital tools, which limits the possibility of formally evaluating convergent validity<sup>5</sup>. Nevertheless, establishing convergent validity is a crucial step to further confirm the construct validity of the UAS. Future research should aim to compare UAS scores with validated measures of related constructs, such as the System Usability Scale (SUS) or Technology Acceptance Model (TAM) based instruments, once appropriate child-adapted Italian versions become available. Conducting such analyses would allow us to determine whether UAS scores converge with theoretically aligned constructs, providing robust empirical support for its validity in assessing children's interactions with digital technologies. Additionally, the current validation was conducted exclusively in educational settings, aware that to enhance the ecological validity and broaden the applicability of the UAS beyond educational settings, it would be valuable to validate the instrument in different contexts, such as clinical and therapeutic environments. We are planning to test the UAS with clinical populations, which will provide important insights into its applicability for assessing digital technologies used in assessment and intervention contexts.

Additionally, while the UAS captures key aspects of usability and acceptability, it may benefit from further refinement and expansion to

include additional dimensions such as long-term engagement and user satisfaction. Moreover, longitudinal studies could provide important insights into how these perceptions evolve over time and with prolonged use of digital technologies. Despite these limitations, the UAS offers a novel and practical tool for evaluating usability and acceptability in children, addressing a gap in the current assessment landscape.

To conclude, our results support the two-factor structure of the UAS, aligning with the theoretical constructs of usability and acceptability. The EFA results support a two-factor structure for the UAS, capturing distinct but related constructs of usability and acceptability. The identified factors are both meaningful and interpretable, reflecting the theoretical underpinnings of the scale. The CFA results provide robust evidence supporting the validity and reliability of the two-factor structure of the UAS, making it a valuable instrument for evaluating digital tools designed for children. The good internal consistency and significant factor loadings indicate that the UAS could be a reliable tool for assessing children's usability and attitudes toward digital tools like MatriKS.

The development and validation of the UAS represent a significant contribution to the field of child-centred digital technology assessment. By providing a reliable and valid measure of usability and acceptability, the UAS helps ensure that digital technologies designed for children are both effective and engaging. This, in turn, can enhance the educational and developmental outcomes associated with the use of these technologies, ultimately contributing to better learning experiences and improved quality of life for young users.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the Ethical Committee for the Psychological Research of the University of Padova, Italy (protocol code E047A9B0520E9732A8365DC72335EE90, date of approval: 8 July 2022). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. This study was conducted in accordance with the Declaration of Helsinki.

## Author contributions

MS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. MB: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. NM: Data curation, Investigation, Validation, Visualization, Writing – review & editing. MO: Data curation, Investigation, Methodology, Visualization, Writing – review & editing. LS: Funding acquisition, Project administration, Supervision, Writing – review & editing. PA:

Writing – review & editing. DC: Conceptualization, Writing – review & editing. AB: Investigation, Writing – review & editing. IP: Investigation, Writing – review & editing. SGa: Writing – review & editing. GB: Funding acquisition, Methodology, Validation, Writing – review & editing. SGI: Conceptualization, Formal analysis, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Italian Ministry of Research and University, PRIN: Progetti di Ricerca di Rilevante Interesse Nazionale - Bando 2020, Protocol n. 20209WKCLL, Project title: Computerized, adaptive, and personalized assessment of executive functions and fluid intelligence.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors SG, MB declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by *Frontiers* with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2026.1702085/full#supplementary-material>

## References

- AlGhannam, B. A., Albustan, S. A., Al-Hassan, A. A., and Albustan, L. A. (2017). Towards a standard Arabic system usability scale: psychometric evaluation using communication disorder app. *Int. J. Hum.-Comput. Interact.* 34, 799–804. doi: 10.1080/10447318.2017.1388099
- Bauw, E., and Markopoulous, P. (2004). A comparison of think-aloud and post-task interview for usability testing with children. In: *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community* (pp. 115–116). New York: Association for Computing Machinery
- Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Int. J. Hum.-Comput. Interact.* 24, 574–594. doi: 10.1080/10447310802205776
- Banker, A., and Lauff, C. (2022) Usability testing with children: history of best practices, comparison of methods & gaps in literature. In: D. Lockton, S. Lenzi, P. Hekkert, A. Oak, J. Sádaba and P. Lloyd. (eds.), *Proceedings of the DRS 2022*: London: Design Research Society
- Barendregt, W., Bekker, M. M., Bouwhuis, D. G., and Bauw, E. (2006). Identifying usability and fun problems in a computer game during first use and after some practice. *Int. J. Hum.-Comput. Stud.* 64, 830–846. doi: 10.1016/j.ijhcs.2006.03.004
- Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: if you feel the need, there is a better alternative to using those negatively worded stems. *Educ. Psychol. Meas.* 60, 361–370. doi: 10.1177/00131640021970592
- Bell, A. (2007). Designing and testing questionnaires for children. *J. Res. Nurs.* 12, 461–469. doi: 10.1177/1744987107079616
- Blažica, B., and Lewis, J. R. (2015). A Slovene translation of the system usability scale: the SUS-SI. *Int. J. Hum.-Comput. Interact.* 31, 112–117. doi: 10.1080/10447318.2014.986634
- Borkowska, A., and Jach, K. (2017) Pre-testing of polish translation of system usability scale (SUS). In: *Proceedings of 37th International Conference on Information Systems Architecture and Technology-ISAT 2016-Part I* (pp. 143–153). Cham: Springer International Publishing
- Borsci, S., Federici, S., and Lauriola, M. (2009). On the dimensionality of the system usability scale: a test of alternative measurement models. *Cogn. Process.* 10, 193–197. doi: 10.1007/s10339-009-0268-9
- Brooke, J. (1996). “SUS: a quick and dirty usability scale,” in *Usability Evaluation in Industry*, eds. P. W. Jordan, B. Thomas, I. McLelland and B. A. Weerdmeester (London: Taylor and Francis), 189–194.
- Brooke, J. (2013). SUS: a retrospective. *J. Usability Stud.* 8, 29–40. doi: 10.5555/2817912.2817913
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema* 20, 872–882
- Byrne, B. M. (2013). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. London: Routledge.
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902\_5
- Costello, A. B., and Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract. Assess. Res. Eval.* 10, 1–9. doi: 10.7275/jyj1-4868
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 319–340. doi: 10.2307/249008
- Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989). Technology acceptance model. *J. Manag. Sci.* 35, 982–1003. doi: 10.1287/mnsc.35.8.982
- de Chiusole, D., Spinoso, M., Anselmi, P., Bacherini, A., Balboni, G., Mazzoni, N., et al. (2024). PsycAssist: a web-based artificial intelligence system designed for adaptive neuropsychological assessment and training. *Brain Sci.* 14:122. doi: 10.3390/brainsci14020122
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., et al. (2007). A psychometric toolbox for testing validity and reliability. *J. Nurs. Scholarsh.* 39, 155–164. doi: 10.1111/j.1547-5069.2007.00161.x
- Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Dev.* 71, 44–56. doi: 10.1111/1467-8624.00117
- Dianat, I., Ghanbari, Z., and Asgharijafarabadi, M. (2014). Psychometric properties of the Persian language version of the system usability scale. *Health Promot. Perspect.* 4, 82–89. doi: 10.5681/hpp.2014.011
- Dizon, G. (2016). Measuring Japanese EFL student perceptions of internet-based tests with the technology acceptance model. *Teach. English. Second. Foreign Lang.* 20:2.
- Doignon, J. P., and Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *Int. J. Man-Machine Stud.* 23, 175–196. doi: 10.1016/s0020-7373(85)80031-6
- Donker, A., and Markopoulos, P. (2002). “A comparison of think-aloud, questionnaires and interviews for testing usability with children,” in *People and Computers XVI-Memorable Yet Invisible: Proceedings of HCI 2002*, eds. J. Finlay, F. D tienne and X. Faulkner (London: Springer), 305–316.
- Downey, L. L. (2007). Group usability testing: evolution in usability techniques. *J. Usability Stud.* 2, 133–144.
- Druin, A. (2002). The role of children in the design of new technology. *Behav. Inf. Technol.* 21, 1–25. doi: 10.1080/01449290110108659
- Dunn, T. J., Baguley, T., and Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* 105, 399–412. doi: 10.1111/bjop.12046
- Emerson, E., Felce, D., and Stancliffe, R. J. (2013). Issues concerning self-report data and population-based data sets involving people with intellectual disabilities. *Intellect. Dev. Disabil.* 51, 333–348. doi: 10.1352/1934-9556-51.5.333
- Fabrigar, L. R., and Wegener, D. T. (2012). *Exploratory Factor Analysis*. Oxford: Oxford University Press.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods* 4, 272–299. doi: 10.1037/1082-989x.4.3.272
- Falmagne, J. C., and Doignon, J. P. (2010). *Learning Spaces: Interdisciplinary Applied Mathematics*. Berlin: Springer Science & Business Media.
- Finlay, W. M., and Antaki, C. (2012). How staff pursue questions to adults with intellectual disabilities. *J. Intellect. Disabil. Res.* 56, 361–370. doi: 10.1111/j.1365-2788.2011.01478.x
- Gronier, G., and Baudet, A. (2021). Psychometric evaluation of the F-SUS: creation and validation of the French version of the system usability scale. *Int. J. Hum.-Comput. Interact.* 37, 1571–1582. doi: 10.1080/10447318.2021.1898828
- Heller, J., and Stefanutti, L. (2024). “Knowledge structures and their competence-based extension,” in *Knowledge Structures: Recent Developments in Theory and Application*, eds. L. Stefanutti and J. Heller (Singapore: World Scientific), 3–26.
- Hourcade, J. P. (2007). Interaction design and children. *Found. Trends Hum.-Comput. Interact.* 1, 277–392. doi: 10.1561/11000000006
- H ysniemi, J., H m l inen, P., and Turkki, L. (2004). Wizard of Oz prototyping of computer vision based action games for children. In *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community* (pp. 27–34). New York: Association for Computing Machinery
- Hvidt, J. C. S., Christensen, L. F., Sibbersen, C., Helweg-J rgensen, S., Hansen, J. P., and Lichtenstein, M. B. (2020). Translation and validation of the system usability scale in a Danish mental health setting using digital technologies in treatment interventions. *Int. J. Hum.-Comput. Interact.* 36, 709–716. doi: 10.1080/10447318.2019.1680922
- ISO (1998). *ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability*. Geneva: ISO.
- Jacoby, S. (2011). *PlayCubes: A Longitudinal Study of Tangible User Interface for Assessment and Intervention of Dynamic Constructional Processes Among Typically Developing Children*. Haifa: University of Haifa.
- Jeffries, R., and Desurvire, H. W. (1992). Usability testing vs. heuristic evaluation: was there a contest? *SIGCHI Bull.* 24, 39–41. doi: 10.1145/142167.142179
- Kantosalo, A., and Riihiaho, S. (2019). Usability testing and feedback collection in a school context: case poetry machine. *Ergon. Des.* 27, 17–23. doi: 10.1177/1064804618787382
- Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: Guilford.
- Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. 5th Edn New York: Guilford.
- Kortum, P., Acemyan, C. Z., and Oswald, F. L. (2021). Is it time to go positive? Assessing the positively worded system usability scale (SUS). *Hum. Factors* 63, 987–998. doi: 10.1177/0018720819881556
- Kramer, J. M., Kielhofner, G., and Smith, E. V. (2010). Validity evidence for the child occupational self assessment. *Am. J. Occup. Ther.* 64, 621–632. doi: 10.5014/ajot.2010.08142
- Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner’s Guide to User Research*. Amsterdam: Elsevier.
- Lee, Y., Kozar, K. A., and Larsen, K. R. (2003). The technology acceptance model: past, present, and future. *Commun. Assoc. Inf. Syst.* 12:50. doi: 10.17705/1CAIS.01250
- Lee, B. C., Yoon, J. O., and Lee, I. (2009). Learners’ acceptance of e-learning in South Korea: theories and results. *Comput. Educ.* 53, 1320–1329. doi: 10.1016/j.compedu.2009.06.014
- Lewis, J. R. (2018). The system usability scale: past, present, and future. *Int. J. Hum. Comput. Interact.* 34, 577–590. doi: 10.1080/10447318.2018.1455307
- Lilley, M., Barker, T., and Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Comput. Educ.* 43, 109–123. doi: 10.1016/j.compedu.2003.12.008
- Markopoulos, P., and Bekker, M. (2003). Interaction design and children. *Interact. Comput.* 15, 141–149. doi: 10.1016/S0953-5438(03)00004-3

- Markopoulos, P., Read, J. C., MacFarlane, S., and Hoysniemi, J. (2008). *Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers*. Amsterdam: Elsevier.
- Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., and Rocha, N. P. (2015). European Portuguese validation of the system usability scale (SUS). *Proc. Comput. Sci.* 67, 293–300. doi: 10.1016/j.procs.2015.09.273
- Marzuki, M. F. M., Yaacob, N. A., and Yaacob, N. M. (2018). Translation, cross-cultural adaptation, and validation of the Malay version of the system usability scale questionnaire for the assessment of mobile apps. *JMIR Hum. Factors* 5:e10308. doi: 10.2196/10308
- Masood, M., and Thigambaram, M. (2015). The usability of mobile applications for pre-schoolers. *Procedia Soc. Behav. Sci.* 197, 1818–1826. doi: 10.1016/j.sbspro.2015.07.241
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah: Lawrence Erlbaum.
- Mosier, C. I. (1947). A critical examination of the concepts of face validity. *Educ. Psychol. Meas.* 7, 191–205. doi: 10.1177/001316444700700201
- Nicolaidis, C., Raymaker, D. M., McDonald, K. E., Lund, E. M., Leotti, S., Kapp, S. K., et al. (2020). Creating accessible survey instruments for use with autistic adults and people with intellectual disability: lessons learned and recommendations. *Autism. Adulthood* 2, 61–76. doi: 10.1089/aut.2019.0074
- Nielsen, J. (1994). *Usability Engineering*. Burlington: Morgan Kaufmann.
- Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 249–256). New York: Association for Computing Machinery
- Nikolaus, S., Bode, C., Taal, E., Vonkeman, H. E., Glas, C. A., and van de Laar, M. A. (2014). Acceptance of new technology: a usability test of a computerized adaptive test for fatigue in rheumatoid arthritis. *JMIR Hum. Factors* 1:e4. doi: 10.2196/humanfactors.3424
- Norman, G. R., and Streiner, D. L. (2008). *Biostatistics: The bare essentials*. PMPH USA (BC Decker).
- Nunnally, J. C. (1978). *Psychometric Theory*. 2nd Edn New York: McGraw.
- Ognjanovic, S., and Ralls, J. (2013). “Don't talk to strangers! Peer tutoring versus active intervention methodologies in interviewing children,” in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, eds. W. E. Mackay, S. A. Brewster, and S. Bødker (New York: Association for Computing Machinery), 2337–2340.
- Perrig, S. A., and von Felten, N. (2025) Development and Validation of a Positively Worded German Version of the System Usability Scale (SUS). *International Journal of Human-Computer Interaction*, 41, 10399–10419. doi: 10.1080/10447318.2024.2434720
- Piaget, J. (1970). *Piaget's Theory*, vol. 1 New York: Wiley.
- Pilotte, W. J., and Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educ. Psychol. Meas.* 50, 603–610. doi: 10.1177/0013164490503016
- Polson, P. G., Lewis, C., Rieman, J., and Wharton, C. (1992). Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *Int. J. Man-Mach. Stud.* 36, 741–773. doi: 10.1016/0020-7373(92)90039-n
- Raita, E., and Oulasvirta, A. (2011). Too good to be bad: favorable product expectations boost subjective usability ratings. *Interact. Comput.* 23, 363–371. doi: 10.1016/j.intcom.2011.04.002
- Ratna, P. A., and Mehra, S. (2015). Exploring the acceptance for e-learning using the technology acceptance model among university students in India. *Int. J. Process Manag. Benchmarking* 5, 194–210. doi: 10.1504/IJPM.2015.068667
- Read, J. C. (2008). Validating the fun toolkit: an instrument for measuring children's opinions of technology. *Cogn. Technol. Work* 10, 119–128. doi: 10.1007/s10111-007-0069-9
- Read, J. C., and MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In: *Proceedings of the 2006 Conference on Interaction Design and Children-IDC'06*. New York: Association for Computing Machinery
- Rubin, J., and Chisnell, D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests*. John Wiley & Sons.
- Salkind, N. J. (2010). “Face validity,” in *Encyclopedia of Research Design*, ed. N. J. Salkind (Thousand Oaks: SAGE), 2455.
- Sauro, J., and Lewis, J. R. (2011) When designing usability questionnaires, does it hurt to be positive? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2215–2224). New York: Association for Computing Machinery
- Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: test of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* 8, 23–74. doi: 10.23668/psycharchives.12784
- Schriesheim, C. A., and Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: the effect on questionnaire validity. *Educ. Psychol. Meas.* 41, 1101–1114. doi: 10.1177/001316448104100420
- Sim, G., MacFarlane, S., and Read, J. (2006). All work and no play: measuring fun, usability, and learning in software for children. *Comput. Educ.* 46, 235–248. doi: 10.1016/j.compedu.2005.11.021
- Sočan, G. (2000). Assessment of reliability when test items are not essentially tau-equivalent. *Adv. Methodol. Stat.* 15, 23–35.
- Soto, C. J., John, O. P., Gosling, S. D., and Potter, J. (2008). The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *J. Pers. Soc. Psychol.* 94, 718–737. doi: 10.1037/0022-3514.94.4.718
- Sterba, S. K., and Pek, J. (2012). Individual influence on model selection. *Psychol. Methods* 17, 582–599. doi: 10.1037/a0029253
- Stewart, T. J., and Frye, A. W. (2004). Investigating the use of negatively phrased survey items in medical education settings: common wisdom or common mistake? *Acad. Med.* 79, S18–S20. doi: 10.1097/00001888-200410001-00006
- Tabachnick, B., and Fidell, L. (2013). *Using Multivariate Statistics*. 6th Edn New York: Harper Collins.
- Teo, T., Lee, C. B., and Chai, C. S. (2008). Understanding pre-service teachers' computer attitudes: applying and extending the technology acceptance model. *J. Comput. Assist. Learn.* 24, 128–143. doi: 10.1111/j.1365-2729.2007.00247.x
- van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Aust. J. Adv. Nurs.* 25, 40–48. doi: 10.37464/2008.254.1817
- van den Haak, M., de Jong, M., and Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behav. Inf. Technol.* 22, 339–351. doi: 10.1080/0044929031000
- Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* 3, 4–70. doi: 10.1177/109442810031002
- Vanette, D. L., and Krosnick, J. A. (2014). “Answering questions: a comparison of survey satisficing and mindlessness,” in *The Wiley Blackwell Handbook of Mindfulness*, eds. A. Ie, C. T. Ngnoumen and E. J. Langer, vol. 1-2 (Chichester: John Wiley & Sons), 312–327.
- Venkatesh, V., and Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decis. Sci.* 39, 273–315. doi: 10.1111/j.1540-5915.2008.00192.x
- Venkatesh, V., and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manag. Sci.* 46, 186–204. doi: 10.1287/mnsc.46.2.186.11926
- Wang, Y., Lei, T., and Liu, X. (2020). Chinese system usability scale: translation, revision, psychological measurement. *Int. J. Hum.-Comput. Interact.* 36, 953–963. doi: 10.1080/10447318.2019.1700644
- Wharton, C., Rieman, J., Lewis, C., and Polson, P. (1994). “The cognitive walkthrough method: a practitioner's guide,” in *Usability Inspection Methods*, eds. J. Nielsen and R. L. Mack (Hoboken: Wiley), 105–140.
- Williams, J. R. (2008). The declaration of Helsinki and public health. *Bull. World Health Organ.* 86, 650–651. doi: 10.2471/BLT.08.050955
- Wong, N., Rindfleisch, A., and Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *J. Consum. Res.* 30, 72–91. doi: 10.1086/374697