
User Experience challenges for designing and evaluating Computer-Based Assessments for children

Florence Kristin Lehnert
University of Luxembourg
Esch-sur-Alzette
florencekristin.lehnert@uni.lu

ABSTRACT

Computer-Based Assessment (CBA), i.e., the use of computers instead of paper & pencil for testing purposes is now increasingly used, both in education and in the workforce. Along with this trend, several issues regarding the usage of computers in assessment can be raised. With respect to CBA, test validity and acceptance appear at stake during interacting with a complex assessment system. For instance, individual differences in computer literacy (i.e. ability to handle technology) might cause different outcomes that are not related to the problem-solving task. Prior investigation has shown that there is a scarcity of research on the User Experience (UX) in the context of CBA, also due to a focus on adult users. This doctoral thesis aims to adapt and develop new evaluation methods from the Human-Computer Interaction (HCI) field, applied in the context of CBA. The contributions will result in the development of best practices guidelines for both research and practitioners by adopting design and evaluation methods drawn from the field of Child-computer Interaction (CCI).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

IDC '19, June 12–15, 2019, Boise, ID, USA

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6690-8/19/06.

<https://doi.org/10.1145/3311927.3328724>

KEYWORDS

User Experience; Computer-Based Assessment; Child-Computer Interaction; User-Centered Design; Scale Design;

INTRODUCTION

During the last decades, the assessment of students has undergone a profound change. Nowadays, Electronic, Technology or Computer-Based Assessment (CBA) is becoming a major contemporary trend, both in education and in the workforce. As the term already implies, CBA refers to the usage of computers instead of paper and pencil for assessment purposes. An example of a recent CBA application in the education field can be found in OECD's "Program for International Student Assessment" (PISA), aiming to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students (see Exhibit 1 for more information).

The development and ever-growing use of CBA in many different contexts has been followed by global computing in modern societies. Beyond the benefits related to the automation of the assessment process (e.g., online scoring, automatic recording, cost, and time reduction), CBA opens new perspectives for innovation in testing and assessment [1, 5, 8, 10]. Examples of those advantages are the construction of innovative item formats [14] such as dynamic or interactive multimedia items which allow for assessing 21st-century skills. However, we should not assess students just because technology makes it easier or as Abraham Kaplan said "give a boy a hammer and everything he meets has to be pounded" [7]. We should carefully examine when and why we can make use of CBA and what the advantages are which a CBA system brings along compared to a pen and paper assessment. In literature, we can find empirical evidence that identical paper-based and computer-based tests will not obtain the same results. Such findings are also referred to as the "test mode effect" [2]. Despite this, paper-based tasks have often been transferred directly onto computers with little concern for performance or user experience implications [14]. Issues regarding the validity and the fairness of CBA for all individuals have been highlighted in a recently published paper [19] and thus should be addressed. The authors show that confounding variables can bias CBA results, several of which are directly related to computer use. For instance, individual differences in computer literacy (i.e. ability to handle technology) might cause individually different outcomes that are not related to the problem-solving task but rather to the idiosyncratically experienced usability of the software environment [19]. This observation is consistent with findings from other researchers [2, 5].

To reduce the impact of Human-Computer Interactions (HCI) issues and to increase the CBA validity, it is necessary to encounter the User Experience of the CBA users. Also in terms of good usability of the interface (e.g. the interface is easy to understand, easy to use, and easy to learn), the user experience can be safeguarded enabling the user to concentrate on the primary task of the assessment [19].

“For the first 12 years, PISA used pen and paper tests because this allowed access to the widest range of learners across the world. However, from the 2015 event, computer-based assessment will be the default mode. The information retrieved from the PISA study plays a major part in the development of educational policies across the world. It allows educationalists to see where the strengths and weaknesses are worldwide and gives them the opportunity to amend education systems accordingly”

Exhibit 1: The Programme for International Student Assessment (PISA).

Moreover, a fair CBA system should take into account many characteristics of the users, for instance, the age, culture, or any other specific needs [16]. Among those characteristics, age is an essential one, in that current research on UX has mainly focused on adult users [6]. In that regard, we can state that children have very particular user characteristics, other than those of an adult user. Therefore, it's necessary to generate further scientific and methodological knowledge about how to design interactive systems for children [12]. Due to the growing interest in a technology-enhanced classroom, along with the increased interest in technological devices in general, the child-computer interaction (CCI) field has emerged over the last years. However, there are still frequently used methods in the UX field which have not been adapted to the child population. For example, Markopoulos & Bekker [12] state that no reports exist of a systematic comparison of usability testing methods (UTMs) focusing on children users. Neither does a systematic effort exist to specify the method and instrumentation of usability testing when test participants are children [12].

Next, regarding the characteristics of children as a target group, other factors should be examined to help understand the underlying psychological needs, motivations, feelings, and attitudes for the acceptance of CBA [11]. During the last years, numerous information technology acceptance models have been developed [3, 18], however, CBA acceptance in the educational field has only been studied recently [4, 11, 17].

RESEARCH OBJECTIVES

This interdisciplinary PhD project aims to improve CBA acceptance for children, by applying methods drawn from the field of HCI and CCI. It also aims at exploring how the design of interactive systems plays a role in the learning process and supports the learning outcomes.

Moreover, the learning environment as part of the user's internal state in relation to the CBA system will be further analyzed. At this juncture, special attention will be drawn to the development and adaption of (current) UX methods to address the target population. Appropriate methods have to be developed and/or adapted to measure other characteristics of the age groups. New developed prototypes will help evaluating the concepts at an early stage and validate our assumptions from a systematic review.

1. Ich bin gut im Fach Rechnen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Ich interessiere mich für das Fach Rechnen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Das Fach Rechnen macht mir Spaß.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Ich habe Angst vor dem Fach Rechnen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Im Fach Rechnen lerne ich schnell.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 1. Example of the Nodding Head Scale of the Luxembourg school monitoring programme (ÉpStan).

OVERVIEW OF STUDIES

Systematic review “State of the art on UX methods for children”

The best suitable UX methods will be selected, supporting psychological needs assessment and design for CBA. A systematic review of UX methods for children between the age 6 to 15 years is currently being conducted and ongoing. This review follows a systematic analysis approach, based on a representative sample of publications in the field of Human-Computer Interaction and UX. The approach is based on a procedure similar to Pettersson, Lachner, Riener, & Butz [15] adapted from the QUOROM statement, which specifies guiding principles on how to conduct meta-analyses including a quantitative data synthesis and a clarifying flow diagram.

UX analysis of the Nodding Head Scale

The “Épreuves Standardisées” (ÉpStan), the Luxembourg school monitoring programme is partner of the Luxembourg Centre for Educational Testing (LUCET) for the largest CBA operation organized in Luxembourg. It consists of written and computer-based tests that create standardized records for mandatory schooling in grade 1, 3, 5, 7 and 9. On a yearly basis, more than 24000 pupils are tested (among whom about 6500 via CBA). The results are used to inform the Government about the impact of local educational policies similar to those of the PISA study. The nodding head scale (see figure 1) is, next to the traditional verbal scale, introduced in the 3rd grade with the underlying assumption that it is more motivating to use faces compared to words or numbers in an assessment setting for children. An ongoing HCI study will validate the findings from the large-scale study by comparing the performance aspects of three different scales (thumbs-up, verbal scale and nodding heads scale) within the same age group (8-9 years) and within the same context (classroom). As part of the HCI study, the children will conduct a think-aloud task to encounter HCI related problems while filling in the scales. To exclude sequence bias, the scales will be randomized per child. A post-interview at the end of the think aloud session will encounter detailed preferences of the children regarding the different scales.

Designing a digital Nodding Head Scale

Based on the results of the previous study and the systematic review, a digital prototype of the nodding heads will be designed, following the User-centered design approach. Suitable methods encountered by the systematic review (e.g. focus groups, usability testing, co-designing workshops), will be used to evaluate the prototype. Within several iterations, the prototype will be adapted and refined (from low to high-fi design) to test the UX design principles and the acceptance of the new scale design among children.

User needs assessment and psychological needs-driven design tools

The best suitable UX methods supporting psychological needs assessment (extracted from the systematic review) will be selected and adapted. In prior research, the HCI research group in Luxembourg developed a user needs assessment tool (“UX cards”) [9], which will be verified for its adequacy in the CBA context with children. Adaptation to participants is not required as the instrument is not self-administered (delivery requires expert support). This tool will generate qualitative data, thus providing additional methodological sophistication.

Adaptation of Evaluation Scales

Self-reported UX scales such as AttrakDiff, UEQ or meCUE as well as current UX context scales may be inadequate for a younger population (wording, format) and lack a precise CBA scope. Acceptance scales already include a subset of UX dimensions (perceived usefulness, perceived ease of use, perceived playfulness); however, they (1) are not adapted to the CBA population properties, (2) are not comprehensive, and (3) only include UX dimensions by coincidence without being rooted in recent UX theory. These scales will thus undergo verification and adaptation for the child population.

IMPLICATIONS

By tackling the above described objectives and studies, we will provide the following three complementary contributions to research needs for the CCI community: Empirical evaluation of the positive impact of UX on CBA validity, acceptance and its role in the learning process; Research and development of new HCI methods (by combining and adapting current methods) for a better integration of UX into CBA. The contributions will finally result in the development of best practices guidelines for both research and practice on the application of CBA for children.

REFERENCES

- [1] Bennett, R. E. (1998). *Reinventing assessment: Speculations on the future of large scale educational testing*. Princeton, NJ: Educational Testing Service, Policy Information Center.
- [2] Clariana, R., & Roy, P. W. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology* (Vol. 33). Retrieved from <https://pdfs.semanticscholar.org/3e8a/ff00cb390786995d0abb473d8233866b3f9e.pdf>
- [3] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319–340.
- [4] Debus, J. C. W., & Lawley, M. (2016). Benefits and drawbacks of computer-based assessment and feedback systems: Student and educator perspectives. *British Journal of Educational Technology*, 47(2), 294-301.
- [5] Farrell, T., & Rushby, N. (2016). Assessment and learning technologies: an overview. *British Journal of Educational Technology*, 47(1), 106-120.

- [6] Fitton, D., Bell, B., Read, J. C., Iversen, O., Little, L., & Horton, M. (2014). Understanding teen UX. *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI EA '14*, 79–82.
- [7] Horowitz, M. J. (1962). Trends in education (report on the annual meeting of the American Educational Research Association held on February 19–21, 1962). *Journal of Medical Education*, 37, 790–793.
- [8] Kyllonen, P. (2008). New constructs, Methods & Directions for Computer-Based Assessment. In Scheuermann, F. & Björnsson, J. (Eds.) *The transition to Computer- Based Assessment. Lessons learned from the PISA 2006 Computer-based Assessment of Science and implications for large scale*. Luxembourg: Office for Official Publications of the European Communities.
- [9] Lallemand, C., Koenig, V., & Gronier, G. (2014). How Relevant is an Expert Evaluation of User Experience based on a Psychological Needs- Driven Approach? In *Proceedings of the 8th Nordic Conference on Human- Computer Interaction*. New York, NY: ACM (pp. 11-20).
- [10] Lee, M.-K. (2008). Computer-Based Assessment in Science (CBAS): Experiences in Korea. In Scheuermann, F. & Björnsson, J. (Eds.) *The transition to Computer-Based Assessment. Lessons learned from the PISA 2006 Computer-based Assessment of Science and implications for large scale*. Luxembourg: Office for Official Publications of the European Communities.
- [11] Maqableh, M., Masa'deh, R.M.T. and Mohammed, A.B. (2015). The Acceptance and Use of Computer Based Assessment in Higher Education. *Journal of Software Engineering and Applications*, 8, 557-574. <http://dx.doi.org/10.4236/jsea.2015.810053>
- [12] Markopoulos, P., & Bekker, M. (2003). On the assessment of usability testing methods for children. *Interacting with Computers*, 15(2 SPEC.), 227–243. [https://doi.org/10.1016/S0953-5438\(03\)00009-2](https://doi.org/10.1016/S0953-5438(03)00009-2)
- [13] Noyes, J.M., & Garland, K.J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51 (9), 1352-1375.
- [14] Parshall, C., Davey, T., & Pashley, P. J. (2002). Innovative Item Types for Computerized Testing. In W.J. van der Linden and C.A.W. Glas (eds.), *Computerized Adaptive Testing: Theory and Practice* Netherlands: Kluwe Academic Publishers.
- [15] Pettersson, I., Lachner, F., Riener, A., & Butz, A. (2018). A Bermuda Triangle? - A Review of Method Application and Triangulation in User Experience Evaluation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 461:1–461:16. <https://doi.org/10.1145/3173574.3174035>
- [16] Reich, K., & Petter, C. (2009). eInclusion, eAccessibility and design for all issues in the context of European computer-based assessment. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 68–73). Luxembourg: Office for Official Publications of the European Communities.
- [17] Terzis, V., & Economides, A. (2011). The acceptance and use of computer based assessment. *Computers & Education*, 56(4), 1032–1044.
- [18] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- [19] Weinerth, K. (2015). How does usability improve computer-based knowledge assessment?. Doctoral thesis, 2015, Luxembourg. Supervisors: Martin R., Koenig V. and Brunner M.