DESIGNING AND DISSEMINATING REVIEW CRITERIA FOR QUALITY OF TABLET APPS IN PRIMARY SCHOOL MATHEMATICS

Ana Donevska-Todorova & Katja Eilerts

Humboldt-Universität zu Berlin; ana.todorova@hu-berlin.de, katja.eilerts@hu-berlin.de

The last two decades of creating digital vis-à-vis analogue resources for learning mathematics have led to a hyper-production that is not sufficiently verified as being fully effective yet. Therefore, further design of novel resources along to sustainability and dissemination of existing ones through re-design appear as crucial issues that need to be tackled further. In particularly, this relates applications for touch devices in primary school mathematics. The scientific significance of this paper is to deepen the understanding of the quality of the tablet apps and design criteria for analysis of their potentials, e.g., related to contents about space and shape, through a design-based research approach as a convenient methodology for their dissemination along to their design. This is principally relevant for pre- and in- service teachers' informed decisions about implementation of apps in own teaching practices, aspects that we empirically investigate and qualitatively analyse through a DBR cycle.

Keywords: criteria for reviewing quality of tablet apps, potentials of touchpad apps, geometry, dissemination, design-based research

INTRODUCTION AND RELEVANCE OF THE THEME

Years of design research have facilitated identification and documentation of potentials of educational materials and digital tools for learning school mathematics. Yet, inspiring learning and teaching approaches involving tablet apps in mathematics classrooms seem still challenging for both research and practice. Researchers have already addressed the issue of sustainability of results of national projects referring mathematics didactical aspects of technology-enhanced learning environments, e.g. Krauthausen, (2012) in Germany. We witness interrupted projects or finished projects without any updates after completion both in national and international contexts. One reason therefore may be the appearance of apps that is exposed to many factors that cannot be controlled (Krawehl, 2010), e.g., the dynamic feature of the on-line market changing on a daily basis. Preventing bringing good practices up the rear might be a stimulating mission for tangible research actions. Sustainability and scale up of already established good practices remain thought provoking and require long-term active engagement of interdisciplinary research and practitioner teams. Research findings that suggest and study examples of the teaching and learning of mathematics with digital technology as well as innovative solutions of scaling and sustaining impacts were addressed in one of the four themes in the Call for papers at the 14th International Conference on Technology in Mathematics Teaching. The aims of this paper are to show a means for designing review measures for quality of apps that could assist teachers in deciding about the usefulness and efficiency of the apps in their own teaching practices whilst to embrace an approach for sub-sequent dissemination of partial outcomes during and after a complete design-based research process. Consequently, it suggests an innovative elucidation of scaling impacts about quality of touchpad apps, e.g. for geometry through pre-service teacher education and in-service teacher continuous professional development program. An empirical part of the study is also offered to serve as an exemplary for the proposed approach.

REVIEW ON LITERATURE

Defining quality of tablet apps for primary school mathematics

The imbalance between quantity and quality or commercial and educational purpose of existing apps on the market has already been registered in literature (e.g. Callaghan & Reich, 2018; Larkin, 2013). What are 'good' learning environments (LE) for mathematics in primary schools in general (Krauthausen, 2012) and which are main ideas for design of quality LE (Wollring, 2009)? How do these criteria for 'good' LE and crucial ideas for their design relate to environments involving tablet apps (Donevska-Todorova & Eilerts, 2019)? In order to make ourselves clear about what do we understand under the term *quality* of tablet app based learning environments for primary school mathematics we refer firstly to mathematics and its didactics, but also to pedagogical and technological concerns. In particular, we relate prominence of the apps primarily to presentation of the mathematical content and clarity/ simplicity of the goals related to curricula, but also to richness of ways for communication, collaboration and cooperation; structure of challenge; quality of feedback, guidance and rewards; connections of different nature; sense-based interactions and functionality. Since attempts for evaluation of the quality of apps are not new in the research of mathematics education and not only that their number, but also their rigor, tends to increase lately, we have used the above *operational definition* in order to take a closer look at theoretically based frameworks and empirical evidence for its analysis.

Reviewing quality of app based learning environments for mathematics

Measures for quality of technological tools for mathematics education according to three types of fidelity: pedagogical, mathematical and cognitive have been suggested by Dick (2008). However, criteria for quality of different kinds of technology enhanced resources for teaching and learning mathematics, e.g. for dynamic geometry software (e.g., by Kimeswenger, 2017; Trgalova, Soury-Lavergne & Jahn, 2011; Trouche et al., 2013) or e-textbooks and digital curriculum resources (Pepin, Choppin, Ruthven & Sinclair, 2017) are not straightforward transferable to other types of resources e.g., apps for touch pads. In particularly, it is worth mentioning that appraisal measures in generative reviews need to be clarified for corresponding specific types of technological tools, e.g., tablet apps that may facilitate the learning of mathematics at a certain level of education. In this regard, there are already studies that analyse quality of apps for primary school mathematics. For instance, numerous apps for developing differing forms of mathematics knowledge: declarative, procedural and conceptual have been reviewed by Larkin (2013). Further evaluation of apps targeted at geometry used a variety of modified evaluative frameworks, e.g., cluster analysis as "a highly versatile and useful methodology that can assist classroom teachers in their initial selection of Geometry apps, as well as providing additional information on pedagogical approaches to support teachers' classroom practices" (Larkin & Milford, 2018a, p. 102). The authors have argued that "this approach provides more specific information regarding how teachers can coordinate the use of various elements of different apps to support mathematical learning beyond what can be achieved using individual apps" (Larkin & Milford, 2018b, p. 14) than the designed instrument by Namukasa, Gadanidis, Sarina, Scucuglia, and Aryee (2016). The instrument for evaluation of apps for upper primary and junior secondary mathematics by Namukasa et al., (2016) bases on four dimensions: the nature of the curriculum addressed in the app, the degree of actions and interactions afforded by the app, the level of interactivity and the quality of the design and graphic features. It uses a different three-level scale in each of the dimensions to guide teachers when selecting apps. We argue that such instruments struggle to assure validity and reliability of the measures. Therefore, our intention in the overall project is not to create a checklist or an isolated and static online questionnaire for assessment of apps, but rather to offer a virtual place where teachers, designers and researchers can meet through sharing own professional expertise in order advancement of the efficiency of the app designs for specific purposes and improving instructional practices.

THEORETICAL BACKGROUND

Mathematical tasks are the ’driving engine’ in mathematics classroom. They have potentials to activate students cognitively and stimulate their engaged participation. Therefore, their design is the core of planning activities and arrangements in particular when they are wealthy in media of use. For these reasons, drawing on the above literature review and previous work of the first author, this paper bases on the theoretical model about mathematical tasks, task-formats and tablet app rich teaching and learning environments (the ‘golden’ rings in Figure 1), and their essential characteristics (the blue puzzles in Figure 1) (Donevska-Todorova, 2019). More precisely, the crucial characteristics are framed in six categories according to the suggested operational definition for the quality of tablet apps for primary school mathematics (in the previous section): (1) mathematical and didactical meaning, (2) communication, cooperation and collaboration, (3) differentiation, (4) feedback and assessment, (5) connections and networking and (6) logistic and technical support (Figure 1). These categories ground a base for establishing and structuring criteria for didactical quality of app-based tasks and learning environments. The accent is firstly set on the mathematical and didactical matters, and hence, this category is emphasised in the model. The next four categories relate to pedagogical aspects whilst the last one to technical features of the apps.

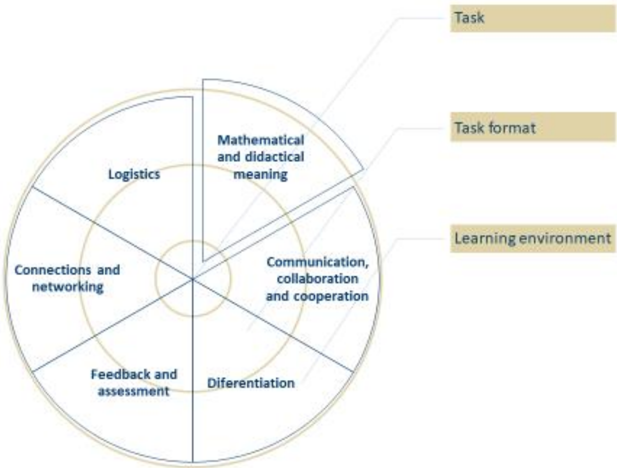

Figure 1. Theoretical model for categories of review criteria for quality of tablet app assisted LE (Donevska-Todorova, 2019, p. 124)

The first category *mathematical and didactical meaning* relates to a structured representation of mathematical concepts and their correctness but also relevance according to content interrelated and process oriented competences defined in national educational standards and curricula. In the empirical part of our study, we have referred to German national standards and curricula for geometry in primary schools in Berlin-Brandenburg. The second category of characteristics *communication, cooperation and collaboration,* tidily refers to ways of interactions between peers (students or teachers), a teacher and students and also devices and users in forms of pair or group activities. The category *differentiation* denotes open possibilities for supporting and providing means for all students, e.g., by offering hints, guidance and adaptive aids along individuals' learning processes. It is closely connected to the category *feedback and assessment,* which aims to show the importance of personalized in-time response and evaluation on given answers or task solutions in regard to both summative and formative forms of assessment. *Networking* means relations between contents in and of the scope of mathematics and natural sciences, between educational haptic and digital materials and networking users. The last category *logistic and technical support* relates user-friendliness of the interface, intuitive handling of the tools, etc. Each of the categories may further be structured into sub-categories by further specification and they are all closely linked to another rather than strictly separated from another.

These categories may also be organized as the three sub-dimensions suggested by Larkin and Milford (2018b), e.g., categories (2), (3) and (4) would be corresponding to the sub-dimension Child-Centred, the categories (1) and (5) to the Learning Design and the category (6) to the Technical Design. Yet, the model (Figure 1) is rather Student-Centred, than Artefact-Cantered, and therefore promoting this feature as three different categories was of importance.

To summarize, the diversity of theoretical frameworks and methodological approaches for identifying reviewing measures for the quality of apps, involve fidelity, cluster analysis, creation of instruments or categorization. These frames and approaches significantly influence the design, re-design and dissemination of apps along to the usage and the evaluation of their quality.

RESEARCH QUESTION AND RESEARCH METHODOLOGY

The overview on literature about the current research state and the theoretical framings have led to a synthesis of our thoughts into a concrete research investigation on the following questions.

What could review criteria for quality of tablet apps for mathematics be, in accordance to the proposed theoretical model (in Figure 1)? How can they be designed and disseminated through teacher training and further professional development?

The relevance of the posed question is two-sided:

(1) helping teachers making informative decisions about usability and efficiency of quality apps and
(2) contributing to sustaining valuable apps through re-design by experts in the research field of mathematics education.

Due to the complexity of the research problem addressing not only potentials of tablet apps based learning environment but also teachers' awareness of their usefulness for practical implementation in classrooms, a single research method does not seem to be appropriate for collecting and analysing data. Moreover, designs of tablet apps change on a daily base. A design-based research (DBR) approach involving different methods at certain of its parts seems to be more suitable. Thus, the DBR intended to design a virtual tool for analysing apps and tasks for apps, enlighten teachers and scale criteria for good app based environments within the first DBR cycle.

Design-based research approaches appear in different forms and under diverse labels through its own historical development (e.g., design experiments, development research, developmental research, design science, design research and design-based research). In the field of mathematics education, there exist DBR models in which there is a clear separation of all phases in two parts, a scientifically one and a practically one (e.g. Prediger, Schnell, & Rösike, 2016). The six phases in the DBR cycle used in this study, an adaptation of the Complete Cycle of Design Research by Kelly, Lesh, & Baek (2008, p. 32) on Figure 2 continuously exchange, beginning with a research phase followed by an applied phase taking place at university. The first phase involved grounded theory approach and qualitative data for creating the theoretical frame (Figure 1) which served as a foundation for the initial design of the review instrument. The second, the fourth and the fifth phases are design experiments, while the third phase tends to measure the validity and reliability of the designed criteria. The last phase aims at arising with a final design that may further initiate a new DBR cycle.

RESULTS AND DISCUSSION

Three paths are relevant for the design of the measures and their dissemination in the suggested approach. Two of them are through the both inner semi-cycles: the first one including phases 1 to 3 and the second one including phases 4 to 6 (Figure 2). The third one is through the complete DBR cycle. They are explained in detail below. An empirical study offering insights about the design and the dissemination of the reviewing measures in praxis through the first inner semi-cycle follows afterwards.

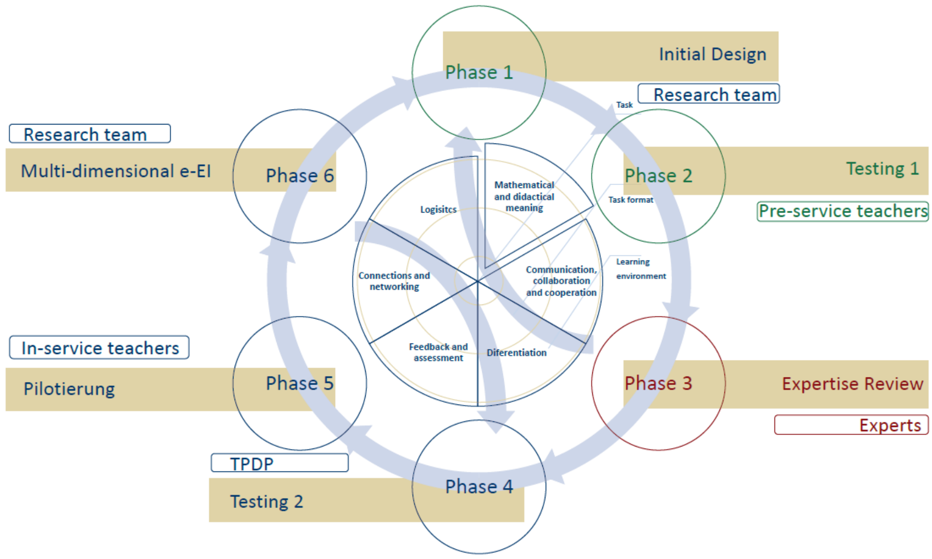

Figure 2. Design-based research cycle of developing reviewing measures evolving from the theoretical model in Figure 1

1. Design and disseminating review measures through the first inner DBR semi-cycle

The first inner DBR cycle begins with Phase 1 that includes the initial design of the reviewing criteria evolving from the categories in the theoretical model situated in the centre of the DBR cycle (Figure 1). It then continues with Phase 2, which comprises a university course for pre-service teachers. This course is a way for dissemination of the reviewing measures along to evaluation of a particular set of apps (empirical part of the study below). The third phase is envisioned as an expertise review of the created initial criteria. After it, certain modification of the design may be required. Therefore, an inner cycle continuing with the beginning phase after phase 3 instead of direct transfer into phase 4 seems much reasonable (therefore made explicitly visible in Figure 2).

2. Design and disseminating review measures through the second inner DBR semi-cycle

Once an expertise examination of the initial criteria is completed and the initial design has being adapted, it may further enter Phase 4. This phase takes account of a larger teacher professional development program. The participants form a smaller sample of in-service teachers than the one involved in the next phase - the pilot trial for the functionality of the reviewing criteria. These two phases allow further re-design and dissemination of the criteria through revising apps in which in-service teachers may play a significant role. This may contribute to development of teacher design capacities due to possibility for reflexion and repetition of this inner cycle.

3. **Design and disseminating review measures through the complete DBR cycle**

After undergoing the whole DBR cycle the design of the measures should be completed. A whole new cycle may begin with a purpose of scaling by adapting the designed instrument for apps that may support the learning of other mathematical contents, e.g. arithmetic. Another reason therefore may be setting a new specific focus on one of the dimensions from the theoretical model (Donevska-Todorova, 2019), e.g., feedback and assessment in tablet app environments for the learning of geometry.

Empirical evidence for the practical implementation

The development of the reviewing criteria for the quality of apps for mathematics education in primary schools has undergone the first two phases of the DBR cycle by now. On the base of the theoretical model (Figure 1) an initial design, containing criteria for apps for geometry organized in the six categories suggested in the model was designed by using the software Lime Survey. However, the criteria do not represent a closed checklist but are created as questions, part of which are open questions asking teachers for own opinion or best practice examples. Further, they are planned as a part of a comprehensive model for valuation having the characteristics of:

- accessibility (without having time and space limitations),
- interactivity (engaged participation by users),
- structuralism (consisting elements are organized in a particular hierarchy with a clear goal),
- multi-disciplinarily (interconnected domains in mathematics, its didactics, pedagogy and media education) and
- sustainability (maintaining changes in a balanced manner to meet needs and aspirations of future users).

Additionally, accompanying elements to the criteria include a repository of apps[1] that were reviewed and documentation like instructions for teachers and students or scientific literature.

The empirical evidence in Phase 2 of the DBR cycle comprises of an undertaken course for pre-service primary school teachers' education. The course was created and implemented by the first author of this paper and took place at the Humboldt-Universität zu Berlin in the winter semester 2018/19. Forty-eighth participants were working in two groups in duration of 90 minutes seminars for 15 weeks. The Moodle platform was used to maintain the organisation of the disseminating process of the designed criteria. The course intended to enlighten teacher students about the rapidly growing amount of apps that do not necessarily possess high quality for mathematics education, offer a possibility for direct investigation of their potentials and limitations and develop sophisticated opinion about an appropriate and meaningful practise of valuable apps. They also aimed to develop teachers' positive attitudes towards usage and needs for further developments. Moreover, in the practical phase of the course, students visited in-service teachers in schools in Berlin and Brandenburg, shared and discussed the criteria for 20 apps for geometry that were suggested in the repository along to their evaluation. The analysis of the results shows the difference between the early criteria identified by the students on the basis of three apps selected from the repository. The most of them were keywords or prompts as "training spatial abilities", "involving different types of tasks", "self-controlling", "intuitively feasible", "combinable", "flexibly applicable", etc. These prompts were later grouped and structured according to the six categories in the theoretical model (Figure 1) after which an increased engagement with the apps and the designed initial criteria took place. Students were working in groups of maximum three participants and the small groups met with at least one in-service teacher. These activities resulted with developed criteria by the students, which were no longer only keywords, but formulations that are more sensible, accurate and clearly fitted a corresponding category in the model.

CONCLUSIONS AND FURTHER RESEARCH

Based on literature review, this paper has arisen with an operational definition about quality of touchpad apps for primary school mathematics. It then presented an overview of existing approaches for studying apps and their values (in section 2) and theoretically framed imperative categories for review criteria in a hypothetical model (in section 3). The discussion gave rise to the question about possible approaches for dissemination of review measures through design and estimation of apps quality. It then argued that the DBR is an appropriate methodology therefore and possible through two of its inner semi-cycles and the complete initial cycle that were described in detail and lead to answering the tackled research question. This has been further justified with a qualitative analysis of empirical data collected during the project and presented in the fifth section.

Several questions arise for further research. Could the designed criteria for quality of tablet apps in elementary geometry be adapted for other areas of primary school mathematics as for example arithmetic? Could that be possible by re-design of some of the criteria and undergoing a semi-cycle or requires starting a new complete design based research cycle? Could the criteria for quality of tablet apps be transferred, disseminated and scaled in other educational areas different from mathematics education by substituting category 1 in the model? These questions may initiate future research concerning effectiveness of tablet apps, learning arrangements and new technology enhanced learning environments.

NOTES

1. Additional information about the repository may be provided by the authors.

REFERENCES

Callaghan, M. N. & Reich, S. M., (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learning, Media and Technology*, *43*(3), 280-293. DOI: 10.1080/17439884.2018.1498355

Dick, T. P. (2008). Fidelity in technological tools for Mathematics education. In G. W. Blume & M. K. Reid (Eds.), *Research on technology and the teaching and learning of Mathematics* (Vol. 2, pp. 333–339). Cases and perspectives Charlotte, N.C.: Information Age Publishing.

Donevska-Todorova, A. (2019). Towards a Theoretical Foundation for Quality Tablet App-Enriched Learning Environments in Primary School Mathematics Education. *International Journal of Technology in Mathematics Education*, *26*(3), 121-129.

Donevska-Todorova, A. & Eilerts, K. (2019). Design of a multi-dimensional instrument for reviewing the quality of apps for elementary geometry. In Jankvist U., T., van den Heuvel-Panhuizen, M., &Veldhuis, M. (Eds.). Proceedings of the *CERME11*. Freudenthal Group & Freudenthal Institute, Utrecht University (accepted).

Kelly, A. E., Lesh, R. A., & Baek, J. Y. (2008). *Handbook of design research methods in education. Innovations in Science, Technology, Engineering and Mathematics Learning and Teaching.* Routledge Taylor & Francis Group, the United States of America.

Kimeswenger, B. (2017, February). Identifying and assessing quality criteria for dynamic mathematics materials on platforms. Proceedings of the *CERME10.*

Krauthausen, G. (2012). *Digitale Medien im Mathematikunterricht der Grundschule.*Springer-Verlag.

Krawehl, F. (2010). Entwicklung eines Evaluationsinstruments zur fachdidaktischen Bewertung von Unterrichtssoftware für das Grundschulalter. In: *Beiträge zum Mathematikunterricht 2010*, München. Gesellschaft für Didaktik der Mathematik GDM.

Larkin, K. (2013). Mathematics Education. Is there an App for that? In: V. Steinle, L. Ball, & C. Bardini (Eds.), *Mathematics Education: Yesterday, Today and Tomorrow (Proceedings of the 36th Annual Conference of the Mathematics Education Research Group of Australasia)*, (pp. 426–433). Melbourne, VIC: MERGA.

Larkin K. & Milford T. (2018a). Using Cluster Analysis to Enhance Student Learning When Using Geometry Mathematics Apps. In: Ball L., Drijvers P., Ladel S., Siller HS., Tabach M., Vale C. (eds) *Uses of Technology in Primary and Secondary Mathematics Education.* ICME-13 Monographs. Springer, Cham.

Larkin, K. & Milford, T. (2018b). Mathematics Apps—Stormy with the Weather Clearing: Using Cluster Analysis to Enhance App Use in Mathematics Classrooms. In: Calder N., Larkin K., Sinclair N. (eds) *Using Mobile Technologies in the Teaching and Learning of Mathematics. Mathematics Education in the Digital Era*, vol 12. Springer, Cham.

Namukasa, I. K., Gadanidis, G., Sarina, V., Scucuglia, S., & Aryee, K. (2016). Selection of apps for teaching difficult mathematics topics: An instrument to evaluate touch-screen tablet and smartphone mathematics apps. In P. S. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 275–300). Cham: Springer International Publishing.

Pepin, B., Choppin, J., Ruthven, K., & Sinclair, N. (2017). Digital curriculum resources in mathematics education: Foundations for change. *ZDM*, *49*(5), 645-661.

Prediger, S., Schnell, S. & Rösike, K.-A. (2016). Design Research with a focus on content-specific professionalization processes: The case of noticing students' potentials. In S. Zehetmeier, B. Rösken-Winter, D. Potari & M. Ribeiro (Hrsg.), *Proceedings of the Third ERME Topic Conference on Mathematics Teaching, Resources and Teacher Professional Development* (pp. 96-105). Berlin: Humboldt-Universität / HAL Archive.

Trgalova, J., Soury-Lavergne, S., & Jahn, A. P. (2011). Quality assessment process for dynamic geometry resources in Intergeo project. *ZDM - International Journal on Mathematics Education, 43*, 337–351.

Trouche, L., Drijvers, P., Gueudet, G., & Sacristan, A. I. (2013). Technology-driven development and policy implications for mathematics education. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 753–790). New York, NY: Springer.

Wollring, B. (2009). Zur Kennzeichnung von Lernumgebungen für den Mathematikunterricht in der Grundschule. In: Peter-Koop, A.; Lilitakis, G.; Spindeler, B. (Eds.).*Lernumgebungen –Ein Weg zum kompetenzorientierten Mathematikunterricht in der Grundschule.*Offenburg: Mildenberger Verlag, pp. 9–23.

Published in: 14th International Conference on Technology in Mathematics Teaching 2019

DOI: 10.17185/duepublico/70786
URN: urn:nbn:de:hbz:464-20191121-172127-9