
Felix Krawehl, Hamburg

June 9, 2010

Abstract

How can mathematical didactical quality of primary school classroom software be assessed? Even by skimming available products defective weaving patterns catch the eye. But how can plausible, valid and relevant quality attribution be carried out? Theoretical and methodological challenges lie in the 'local' limitations of many mathematical didactical postulates, the construction of a theoretically transparent set of relevant criteria and the necessity of integrating implementation affordances. This paper will spread out theoretical and methodological issues and preliminary evaluation model outlines.

Contents

1 Practical problems inherent in classroom software 2
2 Research problem, aims and rationale 3
3 Theoretical framework and available evaluation approaches 4
4 Research activities and preliminary outlines 5
5 Conclusion and Questions 7
1 Practical problems inherent in classroom software

The role of digital media in comparison to traditional didactical media is still marginal. Still, they profit from massive funding initiatives for ‘pedagogical innovation’. They are promoted for the sake of digital literacy and, of course, it’s producing industry. According to latest research (Krützer and Probst, 2006), at least in Germany, classroom software has an important, if not a leading, role among other digital media such as reference software or tools. Classroom software is supposed to incorporate a number of media specific options interesting for the teaching and learning of mathematics. These specific options rely on an appropriate didactical implementation - in the software and in the didactical classroom setting. Both are dubious in quality.

The mathematical didactical ‘weaving patterns’ of classroom software have been criticized since the taking-off of personal computers (Krauthausen, 1992). Among the aspects questioned were drill and rote learning as dominating patterns; lack of curricular validity and naive conceptions of teaching and learning implemented; lack of helpful feedback in terms of subject matter relevance; and misleading suggestions of diagnostic as well as other didactical capabilities. Even a rough skimming through available products easily conveys the impression that this quality – or the lack thereof – have not changed at large in the last two decades (Krauthausen, 2008).

There is little empirical research - at least for Germany - concerning the quality of implementation. Some conditions suggest doing good quality teaching with software might be considered challenging:

- **Concurrence of issues and postulates**: Several reference disciplines as well as classroom management issues compete with mathematical didactical (MD) postulates (e. g. media didactics, general didactics) and may sometimes interfer with best practice thinking of MD. For example, the limited availability of a computer studio and naive media didactical intentions might induce a conception of weekly computer based mathematics lessons, though the software may not be suitable for continuous practising. Lack of MD guidelines (see below) weakens the visibility of it’s claim.

- **Questionable software suggestions**: Many software products emanate questionable suggestions of didactical properties or practices and may mislead teachers in their didactical planning. This is amplified by sometimes misleading or factually incorrect promotional catchphrases of software manufacturers (cf. Krauthausen, 1995). For example, a software whose cover promises diagnostic capabilities and well-directed pedagogical support might be mistakenly instrumented for a corresponding learning setting with best intentions – although the software might in fact implement rote learning suitable for mechanical training in the last training phase at best. The ‘transformative effect’ of the digital media which some media didacts promote as a means to modernize teaching may have an underestimated downside in this respect.

- **Questionable software reviews** in popular pedagogical and ICT-related journals may lead to suboptimal choices and purchases of schools (see also below). These trigger a corresponding development on the side of manufacturers.

- **Lack of md guidelines to good teaching with software**: Whereas most textbooks are accompanied by didactical documentation, qualified didactical documentation is seldom available for software.

Evaluation and quality control may be considered natural approaches to improve the problems mentioned above. In fact, many evaluation instruments for software evaluation exist (see 3). Software reviews are regular columns in many teacher journals. While the plenty may be surprising in the face of the lasting quality problem of software given –, it is not at a second glance. Software reviews in popular teacher (and other) journals – in Germany – are of mixed MD quality and origin. Renowned software award ‘expert’ committees are made up by journalists, computer experts, maybe teachers. As for evaluation instruments of scientific origin, mathematical didactical quality is a white spot in most criteria catalogues. Therefore, the development of an evaluation instrument focussing on MD quality seems a rewarding project.
2 Research problem, aims and rationale

In my research project I develop an evaluation instrument for classroom software for primary schools. It's focus is to give credit for fundamental mathematical didactical (md) quality in order to achieve a better understanding of corresponding software design and teaching with software. Applied in evaluation of classroom software this research may contribute to better visibility of md quality principles, allowing more qualified choice and purchase of software, and in the long run producing economical momentum on software manufacturers. It may also be a tool for formative self-evaluation of developers, thus increasing the likelihood of better classroom software in the future. Finally, better development and choices of software and more knowledge about good teaching may lead to better teaching and learning of mathematics and making best use of the specific options of digital media in this cause. Krauthausen, 2008, mentions the following requirements for such an evaluation instrument:

- imperative of MD focus
- importance of broad scope of MD postulates and applicability on diverse types of classroom software and utilizations
- among desirable outcomes are multifunctionality for several use cases of evaluation, eg: research; reference for software development; reviews for teachers focusing on key aspects (Krauthausen, 2008).

The research method of this project might best be paraphrased by design science sensu Wittmann (Wittmann, 1998). Stufflebeam and Shinkfield accordingly speak of research design in evaluative research as a creative process (Stufflebeam and Shinkfield, 2007, p.66) involving a circular progression between theoretical considerations and implementation testing and interlinking the two. Among central priorities of the intended instrument are the theoretical quality focus (as opposed to an empirical proof of quality e. g. as reflected in learning outcomes) and a focus on mathematical didactical researchers as intended audience and user community. The latter two may trigger questions I'd like to address in advance.

Why attribute quality by theoretical analysis – and not empirical proof via learning outcomes? Implementation is important for the quality of software, as it can support specific properties as well as compensate specific weaknesses. An evaluation model design should therefore integrate aspects of implementation for the sake of valid quality rating. On the other hand, as Clement and Martens, 2000, point out, quality attribution through empirical research on learning outcomes is systematically flawed by uncontrollable context variables and the fragility or agility of learning outcomes in 'real' learning processes. Therefore, effects and their interpretation suffer in terms of reliability and validity. Though an attribution through theoretical analysis has unique challenges, too, it seems more promising in respect to the rationale of clarifying mathematical didactical postulates to focus on the ‘weaving pattern’ and leave out empirical implementation issues.

Why address mathematical didactical researchers as users of an evaluation instrument? This predefinition may seem contradictory, as methodological competency and subject matter expertise as prerequisites for quality attribution may be considered inherent core competencies for researchers anyway. It might seem more logical to address a group with a need for the theoretical backing of an evaluation instrument and a primary interest for evaluation, like teachers deciding about software purchases for schools and districts. On the contrary, the purpose of ‘fundamental’ md quality attribution requires – even with an according evaluation instrument given – expertise for adaption and judgement; and it should not be compromised by possibly conflicting demands of practitioners. The apparent contradiction of this orientation might be resolved by an expectation of a useful tool for the end user – not a surrogate for his expertise.
3 Theoretical framework and available evaluation approaches

Selected definitions and requirements  The intended scope for the evaluand software products includes software for mathematical learning designed for classroom usage. Different kinds of products within this scope are to be addressed, but not reference software (e. g. mathematical dictionaries), tools (e. g. word processing or calculation software) or games. While the boundaries may be fluid, the intention being professionally didactical is the key criterion.

Didactical material in a broader sense has been interpreted differently. A central problem is the relation between software and implementation context. This is relevant for interpretation, as the context can alter a quality estimation. Eg, a good practicing software might make a bad automation software; good software can be used for bad teaching outcomes and vice versa (Tergan, 2000). Following Rezat (Rezat, 2009), software – like text books – can be interpreted as an artefact ‘charged’ with meaning and intention. From this understanding I draw the conclusion that quality attribution relies on prior definition of a probable didactical meaning, based on (maybe fallible) evidence.

Quality is a theoretical construct often used in pedagogical research. Though no common definition can be made out (Wirth, 2005), it’s prescriptive or normative nature can be recorded as theoretical core. This understanding requires a distinction of quality definition and product analysis. Quality description is often put into practice through criteria lists. A clean theoretical approach requires quality criteria to be derived from theoretical postulates and not from features of the evaluand; a transparent description of the underlying quality model; and the relation of criteria to this core model. Thereby coherent quality criteria can be formulated, addressing another meta-criterion for good criteria lists. Applying the quality criteria of qualitative research (Steinke, 2004), validity in terms of theoretical integrity as well as appropriateness for the evaluand is another aspect to consider. The latter implies quality attributions not to be drawn from software properties alone, as they rely on proper classroom implementation to become relevant and to take effect.

I addressed my intended research outcome as ‘evaluation instrument’. Though addressing and referring to MD theory, in terms of methodology evaluation research is the appropriate frame. Accordingly, the standards of evaluation (Joint Committee, 2006) apply as directives for this research. As I will not carry out evaluations myself in the first place, but develope an instrument for doing so, a number of standards only apply indirectly or not at all. In so far as my work addresses theoretical questions, the standards and quality criteria of qualitative research appear as appropriate reference.

Issues in available approaches  We have mentioned evaluation being a natural approach of improving the quality of software. While not specifically addressing mathematical didactical quality and primary schools, there are a broad number of approaches for evaluation of classroom software (also called teaching-, learning-, educational software). A synopsis from 1994 analyzes a dozen of elaborated criteria catalogues alone (Meier, 1994a). Wirth (Wirth, 2005) analyzes and categorizes several general evaluation approaches with many examples.

Software evaluation has progressed not only in quantity, but also in methodological quality. Especially in the evaluation of E-Learning there has been a broad methodological discussion with important results (cf. Wirth, 2005; Baumgartner et al., 2004; Schulmeister, 2003). Some issues relevant for evaluation instrument design are the following:

• Though criteria catalogues should be usable for a range of products, they should contain mechanisms for limiting their scope on the actual demands of the evaluand (Meier, 1994b). This seems to be especially true for mathematical didactical postulates specific for certain subject matters and is common MD research practice, as can be seen applied in the rare available approaches from this side (Selter, 2003; Bokhove and Drijvers, 2010).

• Available evaluation instruments sometimes fail to describe the underlying quality model and combine disparate criteria. This leads to sometimes long, incoherent criteria catalogues compromising usability
and the drawing of plausible conclusions (Wirth, 2005).

- Criteria, standards, indicators and data collection methods should not be mixed for the sake of usability of the resulting evaluation instrument (Wirth, 2005).

The fact that esp. older instruments ignore a number of the principles mentioned above set aside, in our eyes the most prominent research problem is the lack of focus on mathematical didactical criteria. Whether a product complies actual standards in terms of teaching and learning of the subject matter is usually comprised in general questions with little informative value. Available mathematical didactical criteria lists exist for secondary school age only (Bokhove and Drijvers, 2010; Collet et al., 2008), or have methodological limitations (Meissner and Becker-Mrotzek, 1995; Selter, 2003). In spite of this actual representation, we consider the criterion of mathematical didactical quality a key to assessing the overall value of a media and its usage.

4 Research activities and preliminary outlines

The outlines of the generic evaluation process model and the first step may be replenished with commentary on other steps and exemplary application in the discussion in August.

Past and present research activities  As has been said above, the core of my work consists of theoretical design work and cursory testing on a trial basis. Though a progress scheme like “theoretical research – design – evaluation” might be applied as a rule of thumb, naturally (?), progress does not advance in such clear steps in my case. Nevertheless, designing – like writing – has to deal with one issue at a time and I can look back on a number of preliminary design decisions as well as look out on upcoming issues.

By now, I have studied the following theoretical aspects:

- MD discussion on classroom software
- classroom software’s theoretical quality and practical application in teaching and learning;
- available approaches to classroom software evaluation;
- principles and standards of evaluative research, applied for software (‘product’) evaluation.

These theoretical studies have been accompanied by browsing through a selection of available software products, trial application of criteria catalogues and interviewing practitioners using classroom software for their teaching of mathematics. These activities led to first design steps, testing, discussing, reworking and attempts of theoretical founding. By now, a preliminary outline has grown for the whole of the evaluation process and some frameworks of selected steps. Presently, I focus on the development of a theoretically sound, relevant, plausibly coherent and ‘complete’ list of criteria, comprising a central chapter. After that, a preliminary user’s manual seems in reach, allowing exemplary testing of the whole instrument. As a self-evaluation I intend to discuss the usability of the instrument and MD plausibility and relevance of exemplary evaluations on the occasion of a workshop in autumn/winter 2010.
Preliminary design outlines To meet the demands described above, an evaluation model consisting of a generic process and an elaboration of utilizable frameworks for each step are to be developed. The evaluation model should be applicable for different purposes, such as

- formative evaluation of a software design,
- development of a best practice scenario for empirical research of software use,
- theoretical backing up of mathematical didactical documentation,
- sample evaluation of software in teacher education,
- assessment of a given software or parts of it for quality appraisal or comparison.

Each of these use cases requires adaptation. While the final ‘product’ might contain technical advice on conducting different types of evaluation with the instrument, I will focus on the last case mentioned above which can be considered a ‘standard’ evaluation case.

Assuming the evaluation aim has been defined and the evaluand (or coherent parts of it) chosen or identified, these steps comprise the generic evaluation process (Fig.1). The final evaluation reporting step has been left out here, as it varies depending on the specific evaluation aims.

The connection between these steps may best be described in a short walkthrough:

1. Referring to the defined frame (spread out by the dimensions of content and representation, learning necessities and teaching opportunities) (step 1),

2. out of a set of standard categories for good teaching of mathematics relevant criteria are chosen (step 2). These may need further focussing till they comprise mathematical didactical standards for the given frame; also, indicators need to be specified (still step 2).

3. According to these indicators, the evaluand product is checked (step 3), identifying features and lacks and their concrete quality.

4. In step 4, features and lacks are projected into a scenario of ‘ideal’ (relating to the prescribed quality demands) implementation. To each of them, necessary or desirable implementation or compensation within the scenario is outlined. The result is a detailed list of product features and implementation demands.

5. Finally, product and media specific qualities and necessary implementation and compensation actions within a best practice scenario are to be correlated. The questions to be answered as a bottom line are: Do the product qualities outweigh the costs (in terms of necessary implementation necessities)? Might another frame provide a better implementation scenario? Do other products relate better within the same frame?

As the initial framing serves as central reference point for the following steps, it shall be explained in more detail.
Framing quality expectations

Theoretical Aspects  Taking considerations concerning the theoretical construction of MD quality and the nature of software as an artifact into account, definition of the ‘frame’ of expectations, aims and issues is essential for valid quality definition and attribution. For example, we’d expect a software module’s error feedback to work differently in the case of a constructive practicing setting and that of an automatization and speed training. Prior definition of the frame for the evaluation will help focus and clarify expectations and also help produce a coherent quality assessment and judgement. Final quality attribution is necessarily limited to this prior definition.

I propose three dimensions for the framing: subject matter and it’s representation, learning necessities, and teaching opportunities. These dimensions shaped out through triangulation of theoretical sources, esp. available evaluation instruments and their categories, experimental application and theoretical refinement. A confirmation ex post facto might be seen in the congruence with the classic didactical idea of the triangle of subject matter, teacher and learner.

This is rather a heuristic model not intended for more than a plausible approximation to a coherent definition. The definition of a frame is, at the beginning and in the very end, a best guess. This is partly due to the common lack of qualified documentation. Even if there is documentation with didactically relevant information and propositions, there might be still a ‘better’ frame from another point of view. Therefore, the definition of frame should be used as working definition, subject to revision. The only fixed condition is that the final quality assessment refers to one didactical frame and points out resulting limitations. While this delimitation of quality attribution to one defined frame might appear as impractical as it prevents easy comparison of different products, I see it as a valuable information for the audience of MD evaluation reports and future users of the software.

Application  The three dimensions intertwine, but I propose to start the definition with the most obvious, the contents and it’s representations – eg multiplication tables, represented by whatever can be made out by initial skimming of the evaluand software. Next, relating MD ‘micro-’ learning-theories for the given subject matter are to be referred to in order to make propositions to learning necessities addressed in the program module. Picking up the same example, the evaluator would rely on her knowledge about learning of multiplication tables and probably situate exercises with the given materials in a mid-position learning phase with specific requirements. The third proposed definition step concludes with assumptions on teaching opportunities. What general and specific mathematical competencies might be promoted here? In the example, this might include mathematical strategies or problem solving as well as argumentation or mathematical communication skills.

5 Conclusion and Questions

While the theoretical part of my work does not produce new MD theory but combines existing postulates, the attempt to clarify a theoretically sound MD evaluation process itself is the ‘new’ contribution of this project. To remain realistic, the plenty of MD postulates will prevent an evaluation instrument every researcher will be contented by. Also, the rapid technological development might change the premises for such an instrument rather sooner than later. Finally, whether it will be applied at all and whether it can promote the aims of the
rationale, might be a rather theoretical hope. On the other hand, feedback from software purchase decision making teachers gave me the impression that (even) practitioners feel a demand for better software and know-how to distinguish it.

While I expect questions and critical feedback on my proposition at the YERME summer school, if there is space for my questions I would like to ask the following:

- To my surprise, I have found only little international reference literature on my topic. Is that due to my lack of skills or is that it?
- How do you judge the theoretical foundation as for this part? What aspects, references do you miss?
- What would you expect from a qualified meta-evaluation of my approach?

References


