AUTOMATED FEEDBACK AT TASK LEVEL: ERROR ANALYSIS OR WORKED OUT EXAMPLES – WHICH TYPE IS MORE EFFECTIVE?

Guido Pinkernell, Lukas Gulden, Marco Kalz

University of Education Heidelberg; pinkernell@ph-heidelberg

This paper reports on a small-scale quantitative study conducted in a middle school in Germany that compared the effects of two types of feedback on reactivating procedural skills with fractions. Tasks and feedback were implemented in a STACK-based digital learning environment that allowed randomization of numerical and graphical elements of a task as well as automated analysis of student responses to each of the numerical or graphical variations of each task. Due to a small data basis, observations are statistically not verified, but nevertheless point to unexpected results: Especially low achievers seem to benefit more from the error analysis type feedback than from feedback that provided fully worked out solutions. If true, this result suggests that for reactivating and practising skills, error-based feedback is more effective than worked out examples.

Keywords: Automated Feedback, Computer Aided Assessment, Error Analysis, Worked Out Examples

TYPES OF FEEDBACK AT THE TASK LEVEL

“Feedback is information provided by an agent (e.g., teacher, peer, book, parent, self, experience)." (Hattie & Timperley, 2007). In the view of a learner, feedback is good feedback when he or she perceives it as "advice for action" (Ras, Whitelock and Kalz, 2016). One may expect a learner to make even better use of the feedback when it is adapted to individual needs and expectations. Following Ras et al. (2016), the quality of adaptive feedback depends on how the feedback content adjusts to the level of cognition and the personality traits of the learner.

Adaptation to cognitive features and personality traits

Whether feedback is supportive depends on the cognitive disposition of the learner. According to Johnson and Priest (2005), merely communication that an answer is correct or not could be sufficient to prompt the learner to revise and correct his answer, if he or she is an expert. Also, drawing the attention to obvious mistakes in calculation and reasoning can encourage a focussed revision of the solution when the learner has already some relevant knowledge in the area (see Ras et al., 2016). Novices, on the other hand, need feedback of an explanatory and supportive kind, such as scaffolding, which offers specific tips for correcting the presented solutions (Rittle-Johnson and Koedinger 2005). In the case of obvious ignorance of the novice, worked out examples are also conceivable as a basis for further attempts to solve the problem. In terms of performance disposition as a persistent personality trait, Shute (2008) gives similar suggestions for conceptualizing feedback, with timing playing a central role here. Information for low achievers should be made promptly, those for high achievers should be delayed to allow time for revision.

Error analysis vs. worked out examples

Recent research has shown that the use of worked out examples leads to comparably high learning gains (cf. Renkl, 2002). This is explained from the view of cognitive load theory that students can focus on what is important by reducing the need to occupy the learner’s mind with performing single computational steps of the procedure involved (Sweller 1994, Scherrmann, 2016). Worked out examples are suitable especially at the beginning of skill acquisition (Kirschner, Sweller & Clark, 2006). However, while novices profit from worked out examples, this might not hold for experts since the solving procedure shown in the worked out example could interfere with the students' own strategies (Renk & Atkinson 2003).

Making errors is generally regarded as an essential part in the learning process. “An 'error' is a fact or process deviating from a norm, which makes it possible in the first place to discern the correct norm-related fact as an opposite to the erroneous fact or process." (Oser and Hascher, 1996). Thus it seems surprising that in the past there has not been much research on the use of error analysis in feedback (Mory, 2004). With growing availability of computer-based assessment systems (e.g. STACK, Moebius, or Viris) there seems to be a growing interest in research on effects of digital feedback based on a detailed automated analysis of a student’s response. In fact, Livne, Livne and Wight (2007) have shown that while both humans and computers were very good at scoring performance, computers were better in identifying error patterns. “Adaptive feedback information can easily be facilitated within a computer-based instruction environment, where the computer can record and analyze the types of errors being made and give appropriate feedback based upon error types" (Mory, 2004).

RESEARCH INTEREST AND QUESTIONS

The goal of this study is to compare two types of feedback “error analysis” and “worked out examples” with respect to effects on students reactivating and practising their skills in performing simple mathematical operations. For this purpose, the study aims at providing answers to the following four questions:

Q1: Is there a difference between error analysis type and worked out example type feedback with respect to acceptance?

Q2: Is there a difference between error analysis type and worked out example type feedback with respect to performance?

Q3: How does – with respect to each error analysis type and worked out example type feedback – the achievement level correlate with performance?

Our implementation of the error analysis type feedback contains a fully worked out example, too, that becomes accessible only after a delay of 60 sec., and only after the student explicitly decides to view it. The following question seeks to find out whether some students provided with error analysis feedback make extensive use of the worked out examples.

Q4. How does – with respect to the error analysis type feedback – the use of the delayed worked-out example correlate with performance?

THE STUDY

Tasks and feedback were part of the digital learning environment platform of the Heidelberger MatheBrücke (mathebruecke.pinkernell.online), which is Debian based Moodle server with the plugin STACK, a computer algebra aided assessment system (Sangwin, 2013). It allows randomizing mathematical questions as well as analysing student response as to predefined error patterns, which then is being used for error related feedback. Hence, STACK seems a suitable digital learning environment for reactivating and practising mathematical skills that students had an introduction to before.

Tasks

The tasks were for practising simple mathematical operations with fractions. They contain reducing and expanding fractions, adding, subtracting, multiplying and dividing fractions, and switching between graphical and numerical representations of fractions. Each task is randomized, i.e. when the student reloads the task it comes with different numerals or graphics.

Feedback types

Each task also comes with an automated feedback routine that is based on an automatic mathematical analysis of the student's answer. Two type feedbacks were implemented. For describing the feedback content we refer to Narciss´ (2008) classification of feedback components (table 1):

Table 1: Content related classification of feedback components (Narciss, 2008)

Simple components of feedback	Knowlede of performance (KP)
	Knowledge of results (KR)
	Knowledge of the correct result (KCR)
Elaborate components of feedback	Knowledge about task constraints (KTC)
	Knowledge about concepts (KC)
	Knowledge about mistakes (KM)
	Knowledge about how to proceed (KH)
	Knowledge about metacognition (KMC)

The worked-out example type feedback first informs the student about whether the answer is correct or wrong (KR). If wrong, it also tells what the correct solution is (KCR) and gives a model step-by-step solution of the problem (KH), cf. Fig.1.

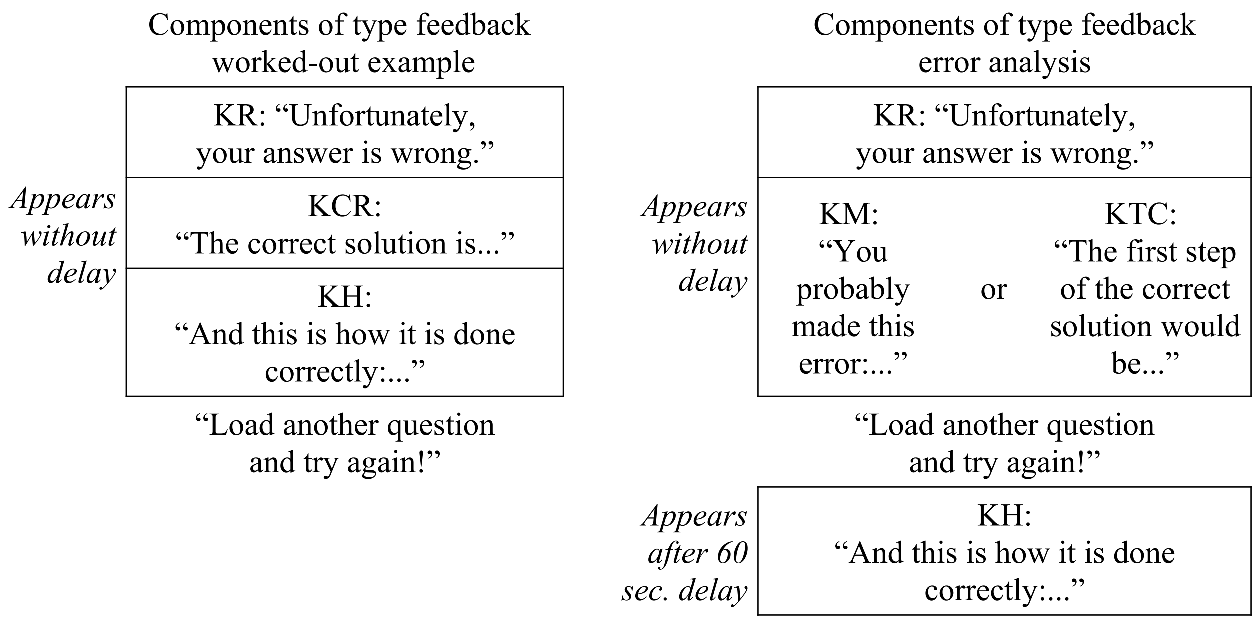

Figure 1: Content related components of the two type feedbacks worked-out example and error analysis

The other feedback type, error analysis, also begins with informing the student that the given answer is correct or wrong (KR). Then, if the student made an error that was previously implemented based on the CAS-based automated error analysis, the student is told what error is made (KM). If the error has not been implemented in the error analysis, the student receives a first step of a suitable procedure (KTC). In both cases, the student then is asked to think about it and try again. Since students of all achieving levels will be part of the error analysis feedback group, we expect some students that rely on a worked-out example. After one minute delay a button will appear which – after click – provides access to a worked-out example.

DESIGN

The treatment was administered to two classes (7th and 8th grade) of a German middle school in the state of Hessia. Each class consisted of 15 pupils between 12 and 14 resp. 13 and 15 years age. All students had been introduced to fractions in grade 6. The students were distributed to two treatment groups such that low and high achievers (identified by their math grade) and were about equally distributed within each groups. In one group the students would be working with tasks with the error analysis type feedback, in the other with tasks with the worked-out example type feedback. One week before treatment the students were given a performance test on fractions which consisted of 8 items that were very similar to the tasks from the treatment. The treatment took place in a computer pool room of their school and lasted about 60 min., after which they filled in an acceptance questionnaire (Cheung & Vogel, 2013, adapted to the digital feedback material of this study as object of acceptance) and, for the error analysis type feedback group only, two use-of-feedback items which asked whether they generally waited for the worked-out example to appear. The posttest consisted of numerical variations of the 8 pretest items and was administered one week after treatment.

RESULTS AND DISCUSSSION

Of the 30 pupils of both classes, 29 took part in pretest, treatment and acceptance as well as use-of-feedback measures. Of these, 4 did not do the posttest, so eventually 25 took part in all testing and treatment measures. Cronbach's alphas for the 10 acceptance items and the two use-of-feedback items were .858 and .574, resp. The data was analysed using non-parametric tests. Except in one case, all tested differences or correlations were not significant. Since all the data seems to point towards a rather unexpected result, we give the details in full.

Q1: Is there a difference between error analysis type and worked-out example type feedback with respect to acceptance?

Fig. 2 shows that there is a very high grade of acceptance of the automated feedback administered in this study. It also indicates that there seems to be a difference between how students rate the two types. Students that worked with type feedback error analysis show a slightly higher grade of acceptance than those that were given the worked-out example type. However, the difference is not significant (Mann-Whitney: $U = 73.5$, $z = 1.35295$, $p = .17702$).

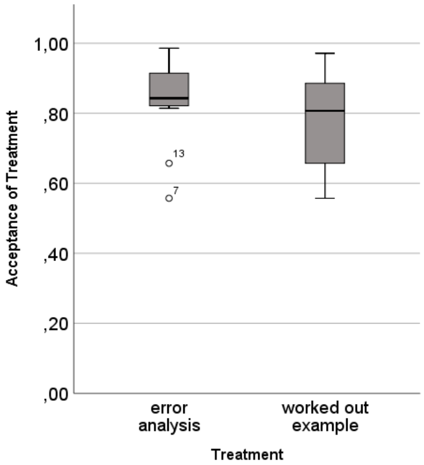

Figure 2: Acceptance rates of digital feedback by students of the two different type feedback treatment groups

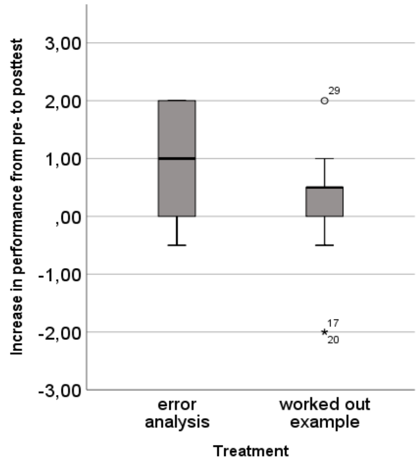

Figure 3: Increase in performance from Pre- to Posttest in the two different type feedback treatment groups

Q2: Is there a difference between error analysis type and worked-out example type feedback with respect to performance?

Fig. 3 shows that students from the treatment group "error analysis" show a slightly higher increase in performance from pre- to posttest. Again, the difference is not significant (Mann-Whitney: U = 52.5, z = 1.35982, p = .17384).

Q3: How does – with respect to each error analysis type and worked-out example type feedback – the achievement level correlate with performance?

While for the worked-out example group the two scales do not seem to correlate (Spearman-Rho: r = .008, p = .98, cf. Fig. 5), there is a high correlation between decreasing math grade and increase in test performance for the error analysis-feedback group (Spearman-Rho: r = .73, p = .007, cf. Fig. 4). It seems that high achievers as a whole do not seem to profit as much from the automatic feedback of both types as the low achievers, which could be explained by a possible ceiling effect due to a pretest that, for some pupils, was too easy. Yet for low achievers the feedback type seems to make a difference, which rather unexpectedly speaks in favour of the error analysis type over the worked-out example type feedback.

Q4. How does – with respect to the error analysis type feedback – the use of the delayed worked-out example correlate with performance?

The error analysis type feedback also provided, after a delay of 60 sec., a worked-out example. After treatment, students of this group were asked whether they generally waited for the worked-out example or started with new tasks right after they were given an error analysis.

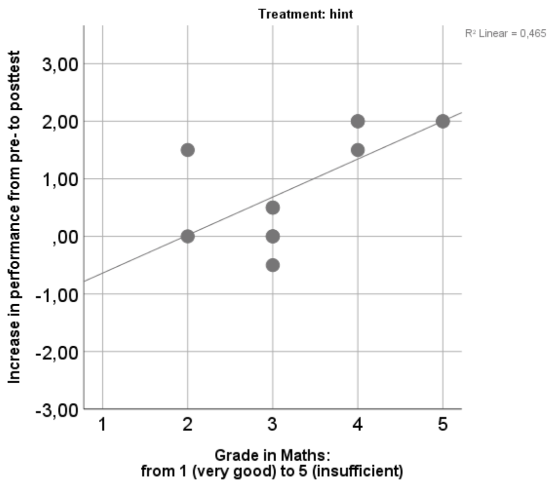

Figure 4: Correlation between increase in performance from Pre- to Posttest and math grades of students from the error analysis type feedback group

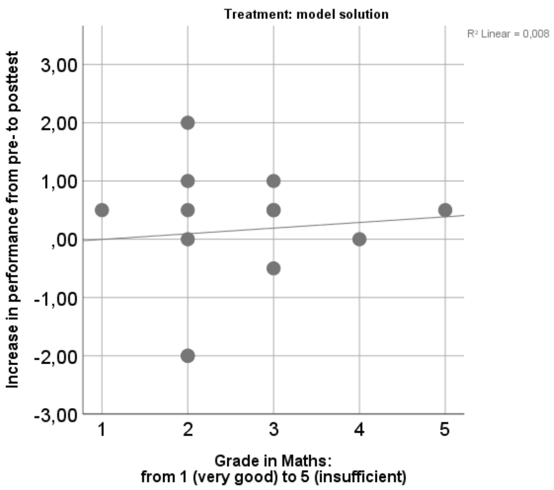

Figure 5: Correlation between increase in performance from Pre- to Posttest and math grades of students from the worked-out example type feedback group

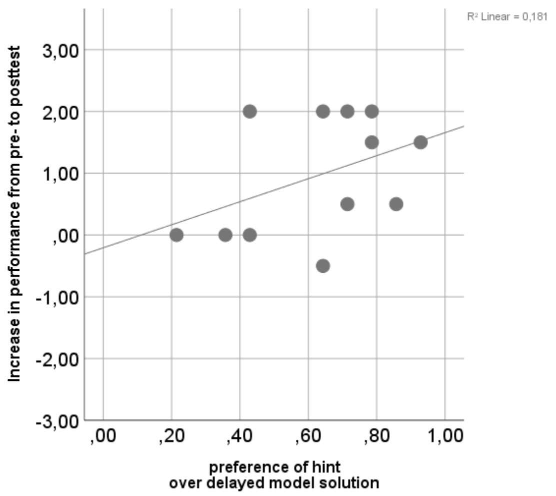

Figure 6: Correlation between increase in performance from Pre- to Posttest and feedback use by students from error analysis type feedback group

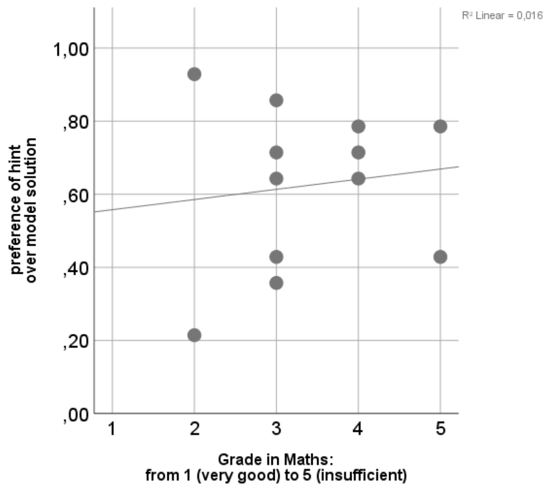

Figure 7: Correlation between feedback use and math grades of students from error analysis type feedback group

Fig. 6 shows that those students who generally did not wait for worked-out examples tend to have a higher performance increase than those who occasionally did (Spearman-Rho: r = .391, p = .208). In fact, there is no correlation between math grade and effective use of the error analysis type feedback (Spearman-Rho: r = .089, p = .784, cf. fig. 7): Within the error analysis feedback group there were high graders as well as low graders who preferred working with the immediately given error analysis or were waiting for the worked out example.

CONCLUSIONS

Due to a very small data base and nearly all observations not being statistically verified, conclusions need to be treated with caution. However, all data seems to point to a rather unexpected result: It seems that low achievers benefit more from error analysis type feedback than from worked-out examples. If this is true, it probably is due to the fact that the treatment was about reactivating skills that had been acquired before. Hence students could not be considered novices any more but experts, for which a worked-out example interfered with the solving strategies they already had in mind, or for which a worked-out example simply did not provide any new information. When students already have some basic understanding of operating with fractions, information reduced to the point of errors made has a better effect on enhancing possibly erroneous or incomplete knowledge. To conclude, it seems that feedback based on an error analysis of the students' responses should be preferred over worked-out examples in the traditional sense in phases of reactivating and practising skills that are already known to students.

REFERENCES

Cheung, R., & Vogel, D. (2013). Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning. *Computers & Education*, *63*, 160–175.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112.

Johnson, C. I., & Priest, H. A. (2005). The feedback principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (S. 449–463). Cambridge: Cambridge University Press.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86.

Livne, N. L., Livne, O.E., & Wight C. A. (2007b). Automated assessment of creative solutions in mathematics through comparative parsing. In A-G. Tan (Ed.), *Creativity: A handbook for teachers* (pp. 399-419). Singapore: World Scientific publishing Co. Pte. Ltd.

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Hrsg.), *Handbook of research on educational communications and technology* (2nd ed, S. 745–784). Mahwah, N.J: Lawrence Erlbaum.

Narciss, S. (2008). Feedback Strategies for Interactive Learning Tasks. In *Handbook of Research on Educational Communications and Technology* (S. 125–144). Mahwah, N.J: Laurence Erlbaum Associates.

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen" Wissens. In W. Althof (Hrsg.), *Fehlerwelten: vom Fehlermachen und Lernen aus Fehlern; Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlaß des 60. Geburtstages von Fritz Oser* (S. 11–42). Opladen: Leske + Budrich.

Ras, E., Whitelock, D., & Kalz, M. (2016). The promise and potential of e--assessment for learning. In P. Reimann, S. Bull, M. Kickmeier-Rust, R. Vatrapu, & B. Wasson (Ed.), *Measuring and visualizing learning in the information-rich classroom* (S. 21–40). New York, NY: Routledge.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529–556.

Renkl, A., & Atkinson, R. K. (2003). Structuring the Transition From Example Study to Problem Solving in Cognitive Skill Acquisition: A Cognitive Load Perspective. *Educational Psychologist*, *38*(1), 15–22.

Rittle-Johnson, B., & Koedinger, K. R. (2005). Designing Knowledge Scaffolds to Support Mathematical Problem Solving. *Cognition and Instruction*, *23*(3), 313–349.

Sangwin, C. J. (2013). *Computer Aided Assessment of Mathematics*. Oxford, UK: Oxford University Press.

Scherrmann, A. (2016). *Lernen mit Lösungsbeispielen im Mathematikunterricht: Eine empirische Untersuchung zur Datenauswertung im Unterricht*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312.

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

DOI: 10.17185/duepublico/70767

URN: urn:nbn:de:hbz:464-20191119-171645-7