# Learning with worked examples – how does it work in a real classroom setting?

Alexandra Scherrmann

Institute of Mathematics and Computer Sciences, University of Education Ludwigsburg, Ludwigsburg, Germany; schermann@ph-ludwigsburg.de

*The learning method "learning with worked examples" was integrated in eleven lesson units about some aspects of descriptive statistics in seven different classrooms at secondary school level. The 194 students (8th graders) were assigned to one of three types of worked examples: with errors, like a cloze and a complete form. The following were measured: whether there are relevant differences in learning success, time spent working on the examples, perceived cognitive load and the acceptance of this way of learning depending on the type of worked examples and the prior mathematical knowledge. The results suggest that in a classroom context the type of worked example is less important than the prior mathematical knowledge, especially for the learning success and acceptance.*

*Keywords: erroneous worked examples, cloze worked examples, statistics classroom.*

## Introduction

Worked examples typically present a task, the solution process in a step-by-step version and the solution itself. In maths classes worked examples typically seem to play a role between the introduction and the exercise of a new mathematical concept and they generally emerge on the board during the explanations given by the teacher (Renkl et al. 2004). However, the learning method "learning with worked examples" has more impact as a range of studies in the field of psychology has shown (see Atkinson, Derry, Renkl, & Wortham, 2000 for an overview). Therefore, as early as 2002, Reiss & Renkl urged for maths classes: "What has to be achieved is 'only' a prolongation and an optimization of the 'example phase'." (Reiss & Renkl, 2002, p.31)

Especially for novice learners in well-structured environments this learning method is more effective than problem-solving practice – with regard to the increase of knowledge in a shorter time-on-task with fewer mistakes. These positive effects have been given a specific name in the research community: They are called the "worked-example effect" (Kirschner, Sweller, & Clark, 2006). But the worked example effect does not exist on its own. Several factors have been identified that influence their effectiveness: These are intra-example features (how the worked examples are designed), inter-example features and individual differences in processing the worked examples, (especially 'self-explanations'). (Atkinson et al., 2000, p.186)

With regard to the intra-example features, presenting different sources of information simultaneously (for example text and graphs) is recommended in order to avoid the split-attention effect (Tarmizi & Sweller, 1988). Also, the worked example should not present redundant information (Kalyuga, Chandler, Tuovinen, & Sweller, 2001) but subgoals and the related solution steps (Catrambone, 1998). A useful inter-example feature is embedding several worked examples with different mathematical contents in the same context or 'cover story', and vice versa – to present various contexts for the same mathematical content, in order to help the learners to focus on structural and not on surface features (Quilici & Mayer, 1996, p.62). But that is not enough. Learners who self-explain the worked example while studying it, were more successful than others (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). The quality not the quantity of self-explanation is crucial (Renkl, 1997).

Several efforts were therefore made to encourage learners to generate self-explanations while studying the worked examples and to process the worked examples in depth. Worked examples like a cloze (Stark, 1999) and others with errors were created (Durkin, 2012; Kopp, Stark, Heitzmann, & Fischer, 2009; Große & Renkl, 2007). Both types of worked example are seen as an indirect way of encouraging self-explanations (see Kopp et al., 2009, p.109).

The extensive evidence for the positive effects of learning with erroneous worked examples or with the cloze worked examples was often furnished by laboratory studies (e.g. Große & Renkl, 2007;) and/or in e-learning settings (e.g. Gerjets, Scheiter, & Catrambone, 2006) and/or with students in colleges or in universities (e.g. Große & Renkl, 2007; Gerjets et al., 2006). Field studies with students were hardly ever carried out (Scherrmann, 2016) and so far no comparison has been made of the three types of worked examples (with errors, like a cloze, complete example). Therefore this study aimed to investigate the effects of working with different types of worked examples, in relation to the prior mathematical knowledge, in a real classroom setting. The following descriptions are part of a more comprehensive study (see Scherrmann, 2016).

## Research question(s) and hypotheses

An experimental field study was carried out in a classroom setting to answer the following question (among others): Depending on the prior mathematical knowledge, what kind of influence do the different types of worked examples (with errors, like a cloze, complete example) have on learning success, the perceived cognitive load, the acceptance of this way of learning and the time-on-task?

Learning success: On the basis of further research (Große & Renkl, 2007) it was hypothesized that learners with high prior knowledge would benefit more from the erroneous worked examples than learners with low prior knowledge (aptitude-treatment-interaction effect). The latter group might be overwhelmed by the task.

Perceived cognitive load: A common explanation for the worked example effect is the cognitive load theory. Obviously, it is difficult to scale the cognitive load as it is a construct relevant to explaining the worked example effect (Brünken et al., 2010). For example, in a study by Stark et al. (2009) in a comparison of complete and erroneous worked examples with medical students there was a small but negative correlation between perceived cognitive load and the learning success.

Acceptance of this way of learning: Under controlled conditions in the laboratory, acceptance seemed to be better with cloze worked examples than with the complete examples (Stark, 1999). Also in the laboratory study carried out by Kopp et al. (2009): Here the erroneous worked examples seemed to have a lower level of acceptance than the complete examples.

Time-on-task: Although the laboratory study with medical university students carried out by Kopp et al. (2009) showed that there was no significant difference in the time-on-task between the erroneous and the complete worked examples, it is assumed that these younger students spend much more time with the erroneous (and the cloze) worked examples than with the complete examples.

# Method

## Sample and design

The study took place in regular mathematics classrooms on the basis of the secondary school mathematics curriculum in Baden-Württemberg, Germany. During a 3-week period with eleven lessons 194 students (8th graders, 91 female and 103 male students) learned about data analysis using various types of worked examples. One type is a complete form of worked example (n=66), another type is similar to a cloze test (n=64) and the third type (n=64) includes errors which were highlighted (because of the results and concerns of Große & Renkl, 2007). First, three parallel groups were formed on the basis of the mathematics grade[1] – this is the prior mathematical knowledge (high, medium, low). Then the students were randomly assigned to one of the types of worked example. The study was divided into two segments. Part A included a pre-, post-, and follow-up test (which is not part of this paper). Part B of the study had two measuring points directly after learning with worked examples.

## Learning environment

The learning environment aimed to give instruction about some aspects of descriptive statistics (mean, median, quartiles), visualizing them in a boxplot and interpreting boxplots. According to Wild & Pfannkuch (1999) it is important during the teaching unit to go through an investigative cycle (problem, plan, data, analysis, conclusions) together with the students (see figure 1).

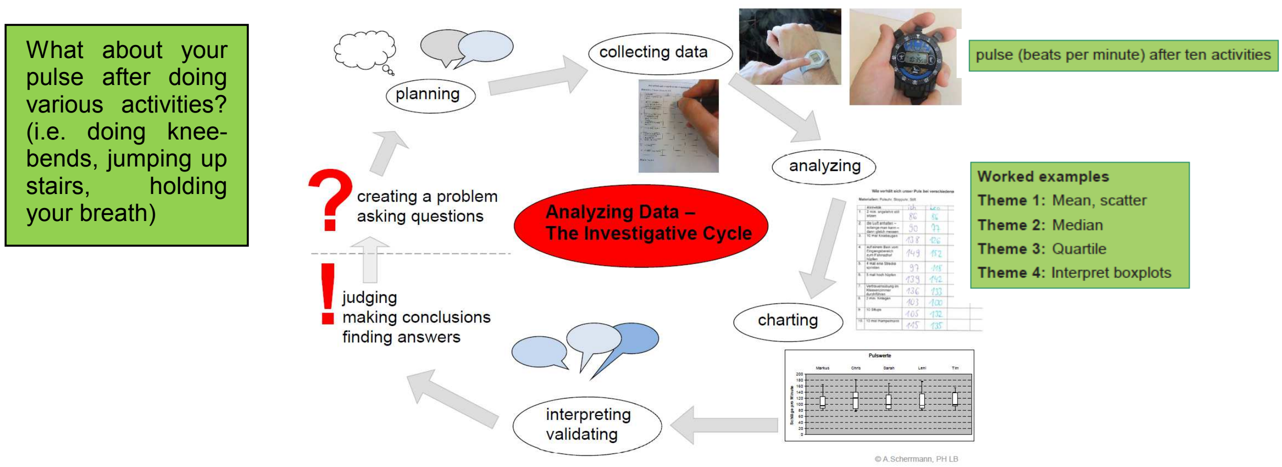


**Figure 1: Worked examples in a statistics classroom**

The teaching aim was embedded in the context of the topic "pulse". The unit was introduced by the teacher and then the students planned the data collection in groupwork (lesson one and two, each

[1] The experimental design was therefore unbalanced. This was taken into consideration in the statistical analyses.

lasting 45 minutes) and measured their pulse during ten activities (lesson three). Then they learned to analyze data (lessons four to eight) and to interpret boxplots (lessons nine and ten) with the help of worked examples. Overall, four worked examples were integrated in this teaching unit.

After studying the actual worked examples the students analyzed their own pulse data. At the end of the unit, the students could interpret and validate their individual pulse data and the pulse data of their classmates related to a specific activity (lesson 11).

**Procedure and implementation of the learning method**

All teacher introductions were nearly standardized by means of a power point presentation and a sequence plan. The first two worked examples about the mean and the median had the special function of introducing the new learning method "worked examples". The students worked with the same type of worked example (with errors, like a cloze, complete form) on their own. For the exercise afterwards and the analysis of their own pulse data they were able to work together with a classmate – but only with a classmate dealing with the same type of worked example. The worked examples were designed with regard to past research (see "Introduction": intra- and inter-example features). Therefore each theme (mean, median, quartile, interpreting boxplots) was embedded in two contexts (rainfall; sending text messages). The context "rainfall" was always presented in a complete form. With the context "sending text messages" the type of the worked example varied: The first type of worked example was with highlighted errors, the students had to improve the errors. The second type of worked examples was like a cloze, the students had to fill in. And the third type of worked example was a complete form, the students had just to read. The time-on-task was measured while learning with worked examples of the theme "Quartile" and the theme "Interpreting boxplots". After each theme, the students had to complete rating scales and a learning test.

**Instruments**

Learning success: This was measured by means of a self-developed issue-specific test with twelve ("Quartiles") or fourteen items ("Interpreting boxplots") respectively. The tests were designed as a class test. The tasks are analogous to the tasks in the worked examples but embedded in another context. Each correct answer received one point, no half points were given.

Insights from the "Quartiles" test:

1.) Determine the following specific values: median, 25% quartile, 75% quartile, distance of the quartiles. [Two datasets with seven or eight disordered values respectively are given, context: measuring pulse in beats per minute when performing different actions].

2.) One dataset has the following specific values [median, quartiles, minimum, maximum]. Which of the following datasets is appropriate to the specific values. Tick off. [Four sorted datasets are given, each with the same minimum and maximum value, each with ten or eleven values.]

Insights from the "Interpreting boxplots" test:

1.) [One horizontal boxplot is given, without a context.] Read out the following specific values: minimum, maximum, median, 25% quartile, 75% quartile.

2.) [Another horizontal boxplot is given, context: cost for 1kg of asparagus.] Which dataset is appropriate to the given boxplot? Tick off. [Four datasets, each with nine or ten values are given.]

3.) [Another vertical boxplot is given, context: cost for a square meter of apartment.] Which statements are appropriate to the boxplot? Tick off "right" or "wrong". a) The spread of the data is €18. b) The bottom half of the values spreads more than the upper half. c) For half of the flats you have to pay at least €9 per square meter). d) 25% of the flats cost more than €12 per square meter.

4.) [Three vertical boxplots are given, context: number of SMS texts sent per week for three grades.] Which statements are appropriate to the boxplots? Tick off. [Four opportunities are given: grade 5 /  grade 8 / grade 10 / no answer can given.] a) 75% of the pupils sent a maximum of 35 SMS texts per week. b) Half of the pupils sent a maximum of 10 SMS texts per week. c) Here the highest number of pupils were asked. d) The values of the central half were most widely spread.

Perceived cognitive load: Perceived cognitive load was assessed by means of a self-developed rating scale with six items based on Hart & Staveland (1988) and Lipowsky et al. (2005), ("I had to think hard today", "It was difficult to keep in mind all the important information in the worked examples", "Today's topic was difficult", "The contents were difficult for me to understand", "Learning with worked examples was stressful", "It was difficult today to distinguish between relevant and irrelevant information") ranging from 1 ("I fully disagree") to 4 ("I fully agree"). Reliability (Cronbach's Alpha) was $\alpha=0.80$.

Acceptance of this way of learning: This was measured by a scale form Stark (1999) with a Likert scale ranging from 1 ("I fully disagree") to 4 ("I fully agree"). The scale involved the following three items ($\alpha=0.76$): "I don't need worked examples to understand of the topic", "What I learned today, I would have learned without worked examples", "Learning with worked examples is too time-consuming for me".

Time-on-task: This was was defined as the time between all beginning together and individual completion of studying the worked examples (it ranged between 7 and 47 minutes).

### Statistical analyses

For the 3×3·2 split-plot factorial design with the two between-subject factors "type of worked example" (erroneous/cloze/complete form) and "prior mathematical knowledge" (low/medium/high) and the inner-subject factor "measuring point" together with multiple dependent variables, MANOVA is an appropriate computing method. An alpha level of 0.05 was used. Effect sizes was calculated with partial $\eta_p^2$. The values are orientated to Cohen's conventions (1988): 0.01 was labelled as weak, about 0.06 as medium and effects higher than 0.14 as strong.

## Results

### Effects of type of worked example on perceived cognitive load and time-on-task

No significant effect was found of the "type of worked example" ($V = 0.36$; $F(10, 216) = 4.75$; $p < 0.001$; $\eta_p^2 = 0.18$; great effect),) on the variable "perceived cognitive load" ($F(2, 111) = 3.46$; $p = 0.035$; $\eta_p^2 = 0.06$; medium effect) and the "time-on-task" ($F(2, 111) = 18.22$; $p < 0,001$; $\eta_p^2 = 0.25$; great effect). Students assigned "erroneous worked examples" showed the lowest perceived cognitive load, those assigned "complete worked example" the highest (see Table 1). Simultaneously the latter group required the shortest time-on-task with about 18 minutes. In contrast, compleing the cloze needed about 25 minutes. The type of worked example showed no statistical effect on the acceptance of this way of learning and on the learning success.

| TYPE OF WORKED EXAMPLE | perceived cognitive load | | time on task | |
|---|---|---|---|---|
| | M | SE | M (minutes) | SE |
| **complete form** | 2,16 | 0,08 | 17,85 | 0,88 |
| **cloze** | 2,03 | 0,08 | 25,33 | 0,88 |
| **erroneous** | 1,88 | 0,07 | 20,85 | 0,85 |

notes: M = mean; SE = standard error

**Table 1: Perceived cognitive load and time-on-task dependent on the type of worked examples (Scherrmann, 2016, p.278)**

### Effects of prior mathematical knowledge on learning success, the acceptance of this way of learning and the perceived cognitive load

In the measurement directly after learning with worked examples, the prior mathematical knowledge also proved to have a significant effect ($V = 0.32$; $F(10, 216) = 4.17$; $p < 0.001$; $\eta_p^2 = 0.16$; great effect) on the following dependent variables: learning success ($F(2, 111) = 15.01$; $p < 0.001$; $\eta_p^2 = 0.21$; great effect), the acceptance of this way of learning ($F(2, 111) = 5.35$; $p = 0.006$; $\eta_p^2 = 0.09$; medium effect) and the perceived cognitive load ($F(2, 111) = 5.19$; $p = 0.007$; $\eta_p^2 = 0.09$; medium effect) (see table 2). Therefore the prior mathematical knowledge was not relevant for the time-on-task.

| PRIOR MATHEMATICAL KNOWLEDGE | Issue-specific learning success | | perceived cognitive load | | acceptance of this way of learning | |
|---|---|---|---|---|---|---|
| | M | SE | M | SE | M | SE |
| **high** | 11,29 | 0,30 | 1,83 | 0,08 | 2,47 | 0,11 |
| **medium** | 9,70 | 0,29 | 2,07 | 0,07 | 2,95 | 0,11 |
| **low** | 9,05 | 0,29 | 2,17 | 0,08 | 2,85 | 0,11 |

**Table 2: Issue-specific learning success, perceived cognitive load and acceptance of this way of learning dependent on prior mathematical knowledge (Scherrmann, 2016, p.280)**

As expected, Table 2 shows that the students with high prior mathematical knowledge achieved the highest learning success. This group also perceived the lowest cognitive load but also the lowest acceptance of this way of learning. In contrast, the students with low prior mathematical knowledge felt the highest cognitive load.

## Discussion

The prior mathematical knowledge predicted the learning success and the affective variables as expected. Learners with high prior knowledge learned more successfully and showed a lower level of perceived cognitive load than learners with poor prior knowledge.In particular, no interaction effect was found between the “type of worked example” and the “prior mathematical knowledge”, which was hypothesized with further laboratory studies (Große & Renkl, 2007; Stark, 1999). Therefore, no general recommendation can be given as to which type of worked example is “best” in a statistical classroom setting. But two comments are to the point: The perceived cognitive load is lowest with the erroneous worked example and highest with the complete form. It seems to be “easier” to improve (highlighted) mistakes than only reading text and tasks in the worked examples. Also, the learners with high prior mathematical knowledge showed the lowest acceptance of this way of learning. That is reasonable in view of the expertise-reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). The best acceptance was observed in learners with a medium level of prior mathematical knowledge (only insignificantly lower for learners with poor prior knowledge). Maybe the learning method “worked examples” suits these learners best.

Further research should be conducted to replicate these findings and could take a look at other mathematical topics.

## References

Atkinson, R. K., Derry, S., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. Review of Educational Research, 70, 181–214.

Brünken, R.; Seufert, T. & Paas, F. (2010). Measuring cognitive load. In: Plass, J.; Moreno, R. & Brünken, R. (Ed.), Cognitive load theory. Cambridge: Cambridge University Press, 181–202

Catrambone, R. (1998). The subgoal learning model: creating better examples so that students can solve novel problems. Journal of Experimental Psychology, *127*(4), 355–376.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145–182.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd). Hillsdale, N.J: L. Erlbaum Associates.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes. Learning and Instruction, 17, 612–634.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? Learning and Instruction, 16, 104–121.

Durkin, K. (2012). The effectiveness of incorrect examples and comparison when learning about decimal magnitude (Dissertation). Graduate School of Vanderbilt University, Nashville, Tennessee. Retrieved from http://etd.library.vanderbilt.edu/available/etd-04022012-004633/unrestricted/Durkin_Dissertation.pdf

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp.139–183). Amsterdam, New York: Elsevier Science Pub. Co.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. Journal of Educational Psychology, 93, 579–588.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. Educational Psychologist, *8*(1), 23–31.

Kirschner, P., Sweller, J., & Clark, R. (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, 75–86.

Kopp, V., Stark, R., Heitzmann, N., & Fischer, M. R. (2009). Self-regulated learning with case-based worked examples: effects of errors. Evaluation & Research in Education, 22, 107–119.

Lipowsky, F., Rakoczy, K., Buff, A., & Klieme, E. (2005). Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie “Unterrichtsqualität, Lernverhalten und mathematisches Verständnis”: Teil 1 Befragungsinstrumente. Frankfurt am Main: DIPF.

Quilici, J., & Mayer, R. (1996). Role of examples in how students learn to categorize statistics word problems. Journal of Educational Psychology, 88, 144–161.

Reiss, K.; & Renkl, A. (2002). Learning to prove: The idea of heuristic examples. Zentralblatt für Didaktik der Mathematik, 34(1), 29–35.

Renkl, A.; Schworm, S. & Hilbert, T. (2004). Lernen aus Lösungsbeispielen: Eine effektive, aber kaum genutzte Möglichkeit, Unterricht zu gestalten. In: Doll, J. & Prenzel, M. (Ed.), Bildungsqualität von Schule. Münster: Waxmann, 77–92.

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. Cognitive Science, 21(1), 1–29.

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

Stark, R. (1999). Lernen mit Lösungsbeispielen: Einfluss unvollständiger Lösungsbeispiele auf Beispielelaboration, Lernerfolg und Motivation. Göttingen: Hogrefe.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. Journal of Educational Psychology, 80, 424–436.

Wild, C., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. International Statistical Review, 67(3), 223–248.