Certainty-based marking as feedback in the context of formative assessment in mathematics lessons at school

Joerg Zender[1] and Martina Geisen[2]

[1]Private School, Frankfurt am Main, Germany; joerg@zender.xyz

[2]University of Potsdam, Germany; martina.geisen@uni-potsdam.de

Self-reflection is an essential skill needed to adjust one's learning process. However, training is necessary to develop this skill. Certainty-based marking is presented as an easy-to-integrate method of formative assessment that strengthens predictive accuracy about one's answers to mathematics problems. Knowing better about one's knowledge might lead to better self-reflection. An explorative study with German 5th graders is presented here, in which the influence of certainty-based marking in formative assessment on pupils' mathematical self-reflection during mathematics lessons is examined exploratory. Initial findings show that the certainty and the accuracy with which the pupils state their knowledge rise over time.

Keywords: Formative assessment, certainty-based marking, self-reflection, middle school, feedback.

Introduction

Self-reflection is a cognitive process involving conscious observation of one's thoughts, feelings, and actions and is essential for understanding individual learning, e.g. for mathematics lessons at school. In dealing with heterogeneity, various forms of differentiation are used in mathematics lessons in which, among other methods, learners choose from the options provided (e. g. Hußmann & Prediger, 2007). However, this can be a challenge for students. They must first learn to assess themselves correctly and then make a selection based on this to control their learning and take responsibility for it in the long term. So, promoting students' self-reflection can improve teaching and learning outcomes. One possibility for training students' self-reflection could be the method of certainty-based marking (CBM) in the context of formative assessment. CBM is a testing format that requires students to express a degree of certainty in their responses which will be considered for grading.

This article presents the initial findings of an ongoing exploratory study that investigates how certainty-based marking in formative assessment influences the self-reflection of 5th-grade learners regarding mathematics at school. In the following sections, certainty-based marking is first explained to address its potential in the context of formative assessment and the need for research in this area. Then, the methodological approach is described, initial results are presented, and finally, the results are reflected upon.

Certainty-based marking as feedback in the context of formative assessment

Studies on feedback in the context of formative assessment were carried out by Brensing et al. (2021). The results show that combined formative assessment and feedback can be a powerful tool for fostering mathematical learning. In their study on a first-semester course of mathematics for engineering, counselling after a formative assessment leads to better results, and even weak students with low mathematical knowledge could come close to the mathematical performance of the middle group. However, the effect did not reach so far that weaker students reached the high-performing

students. The feedback was given in individual counselling sessions, which was quite work-intensive and time-consuming, but it helped the students to adjust their learning and perform better. Following Kruger and Dunning (1999), in these counselling sessions, it was revealed that the weak students did not consider themselves weak. But they thought they would perform much better than they did. They first needed a realistic self-reflection to help them understand what they did not understand to start an efficient learning process.

Certainty-based marking (CBM) is an assessment approach developed and first used by Gardner-Medwin (1995) for teaching physiology, medicine, and maths at University College London. The genesis of CBM can be traced back to the desire for a more nuanced and comprehensive assessment methodology that goes beyond the limitations of traditional grading systems (ibid.). Traditional grading systems often provide a single numerical score for a student's solution to a problem, leaving little room for insight into the students' thought process or the level of confidence in their answer (ibid.). So, instead of assigning a fixed score to a solution, CBM allows students to express their confidence in their answers. Students typically provide the solution to a problem and add a confidence rating, such as high, medium, or low, influencing the grading (e. g. see Table 2). Certainty-based marking encourages students to think critically about their responses and reflect on their understanding of the material. This approach recognizes that learning is a dynamic process, and students may possess varying degrees of confidence even when their answers are correct. By incorporating certainty levels into assessment, educators gain insights into students' self-awareness and understanding of the subject matter.

Later, Yuen-Reed and Reed (2015) simplified the confidence rating to two confidence levels and grades without malus points in contrast to Gardner-Medwin (1995). The two options of "unsure" and "very unsure" are combined into just one option of “unsure” because Yuen-Reed and Reed (2015) argue that this distinction is very subjective.

Today, studies on various facets of CBM are available. It has been used in assessment but more in self-assessment and summative assessments (Gardner-Medwin, 2019). In Germany, CBM was, for example, conducted at the RheinMain University of Applied Sciences in a mathematics course with 43 third-semester students in the Department of Engineering (Kanzinger & Gehrig, 2022). Four tests were conducted, each with five subject-specific single-choice questions and CBM. Results were sent to the students at the end of the day the tests had been conducted. An online survey among the students as an evaluation of CBM took place at the end of the semester. Results indicate that CBM could promote the students' ability to self-reflect on tests during the semester. Almost all students stated that CBM helps them to reflect on their learning. There was no control group in this study (ibid.). Following these ideas, integrating formative assessment with certainty-based marking could represent a powerful and nuanced approach to understanding and enhancing student learning, particularly in disciplines like mathematics. It has been tried in medicine (Hendriks et al., 2019) but only in one assessment inside a series of assessments and in veterinary (Valero & Cárdenas, 2017). Foster (2016) used CBM in mathematics in school and investigated how pupils deal with this approach and respond. For this purpose, a ten-task questionnaire on negative numbers and a certainty grading with ten steps was carried out with different classes from age 10 to age 13. Based on this test, Foster (2016) stated that the pupils were well-calibrated but demanded to repeat the test since they then understood their

certainty better. According to Foster (2016), research is needed into what effects will be there if CBM is done more often.

Formative assessment takes place parallel to the learning unit or course to support the learning process and to improve individual learning (cf. Geisen & Zender, 2023; see also Brookhart, 2010; Cizek, 2010; Maier, 2010; Gikandi et al., 2011) and enables teachers to adapt learning opportunities to the needs of the respective learners (cf. Black & Wiliam, 2009). Introducing certainty-based marking within formative assessment could take this process a step further. Students could receive feedback on their answers and be prompted to self-reflect and express their confidence levels in their answers, which they also receive feedback to. This dual-layered evaluation could acknowledge the dynamic nature of learning and encourage metacognition – the ability to think about one's thinking. The integration of formative assessment with certainty-based marking could cultivate a learning environment that is responsive, adaptive, and focused on continuous improvement. On the one hand, students could become active participants in their learning journey. On the other hand, educators could receive valuable data, not only one-dimensional, on how much the students did correct, but also a second dimension, which is how certain they are. This leads to new insights if a topic remained unclear for the students, although they did perform well. Alternatively, students could be very confident in a topic but answer the questions wrong, which can uncover misconceptions. Based on the data, educators can tailor their instructional strategies to the different types in this now two-dimensional outcome.

The literature shows that CBM is used in the university context (and mainly in medicine) but not much in school so far. It is also not commonly known or used in Germany. Therefore, the study presented below can be seen as a starting point for using CBM in the context of formative assessment in school mathematics lessons.

A Weekly CBM test in mathematics lessons - an exploratory study in fifth grade

Objectives

CBM is rare and plays a minor role in assessment. Nevertheless, the wide range of Gardner-Medwins publications (e. g. 2019) show a promising use of CBM in self-assessments. However, there is a need for research concerning CBM in the context of formative assessment. Furthermore, self-reflection should already be used at school. Using CBM in formative assessment could train pupils at school to reflect on their knowledge and improve on this reflection, a necessary skill for individual learning. Therefore, self-reflection should be initiated early in a learner's biography. However, there is still a need for research in this area. An exploratory study is therefore being conducted to investigate the influence of CBM in formative assessment on middle school pupils' mathematical self-reflection.

Method and sample

To measure the influence of CBM, an experiment fifth-grade class was chosen from a German grammar school. The ten- and eleven-year-old pupils had six hours of mathematics lessons each week, in which the last lesson of each regular week (excluding vacations, weeks in which there were reports and in which a class test was written) was used for a short test on the content of the previous

lessons in that week (for an overview see Table 1). By the time this article was written, eight tests had been set.

Instead of measuring self-reflection directly, the accuracy of the confidence in the answers was measured by how often a student was correct when being sure about the answer or how often a student was not correct when unsure about the answer.

Table 1: Overview of weekly tests in Grade 5

Week	1+2	3 – 7	8 – 11	12 – 14
Content	Repetition primary school	Numbers	Divisibility	Measurement and Units
Test	1	2, 3, 4	5, 6	7, 8

The tests were conducted weekly and contained eight open questions (no multiple choice). Additionally, they had a certainty question after each mathematical problem about how sure the students were about their solution. There were two checkboxes: "I am ☐ sure ☐ not sure.". Based on the concept of Yuen-Reed and Reed (2015), the certainty levels were reduced from three to two to make it simpler for the pupils because of their age. The rating was 2 points for a correct and sure answer, -2 points if the answer was wrong but sure, 1 point for a right but unsure answer, and 0 points if the answer was wrong and unsure. If no checkbox was used, 0 points were assigned regardless of whether the answer was correct or wrong (see Table 2).

Table 2: CBM points scheme (based on Yuen-Reed & Reed, 2015)

	sure	unsure	nothing
correct	2	1	0
incorrect	-2	0	0

So, it is always better to state at least to be unsure than not using the checkbox. Consequently, the pupils have used the checkboxes. Also, in this system, it is easy to understand that it is better to be honest about the certainty of your answer. The tests were conducted in the morning, corrected and returned to the pupils in the afternoon of the same day.

A class of fifth graders from a full-time German grammar school was chosen for a first exploratory study about using CBM within formative assessment in school. The class had n=15 pupils aged 10 and 11 (six male and nine female). The pupils came from the same elementary school, and their respective performance levels in math lessons were heterogeneous.

Empirical analysis and initial findings

Regarding accuracy, the class test results were investigated in two ways: First, by how many correct answers the pupils gave and if they were sure about this, and second, by how many incorrect answers the pupils gave and if they were unsure about this. The results are displayed in Table 4.

Table 4: Pupils' prospective validity in percent

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8
sure and correct	67%	70%	66%	59%	83%	86%	75%	87%
unsure and incorrect	52%	63%	68%	66%	85%	67%	80%	64%
mean	59%	67%	67%	62%	84%	77%	78%	76%

For the first four tests, it looked pretty much the same, but then something changed and the pupils started to rank higher. The diagrams in Figure 1 (see next page) show each pupil represented by one dot. The dot is positioned on the x-axis by how many points the pupil got in the test and on the y-axis by how many correct answers were given by this pupil. A perfect accuracy would mean, that the pupil is on the straight line with gradient ½. As the diagrams show, the points are moving towards that line over time, especially after the fourth test.

Discussion and conclusions

This paper deals with using CBM within formative assessment in school mathematics lessons, which could strengthen the prospective validity of pupils' into their answers to mathematical problems. Knowing more about one's expertise may lead to improved self-reflection, essential for adjusting one's learning process. It is, therefore, necessary to clarify the influence of CBM in formative assessment on the mathematical self-reflection of pupils at school. Additionally, research is needed into the question of what effects there will be if CBM is done more often (Foster, 2016). As a starting point in order to answer these questions, we gathered empirical data by testing one class of 5^{th} graders of a German grammar school with CBM tests almost every week in a school year. By the time this article was written, eight tests had been set.

In terms of accuracy, the results are promising, but they need to be treated carefully since the sample size was tiny. After the first four tests with nearly the same prospective validity, the rate suddenly went up. Pupils get better at telling what they have learned and what not. As stated in the methods section, the self-reflection itself was not measured, but a better accuracy could speak in favour of a better self-reflection. At least, one must know about one's knowledge as a basis of self-reflection. If the pupils then used their knowledge to learn and practice precisely what they had identified as their weaknesses before, this would be a step towards self-regulated learning. However, there has been no further investigation about the consequences of the test results for the pupils so far.

The presented exploratory study is a starting point and is still ongoing, so the findings must be interpreted cautiously. In addition, these findings relate exclusively to one experimental group, so to a small sample. Therefore, the study will initially be continued in the current groups and then expanded to include further groups so that future results will be based on a broader data base.

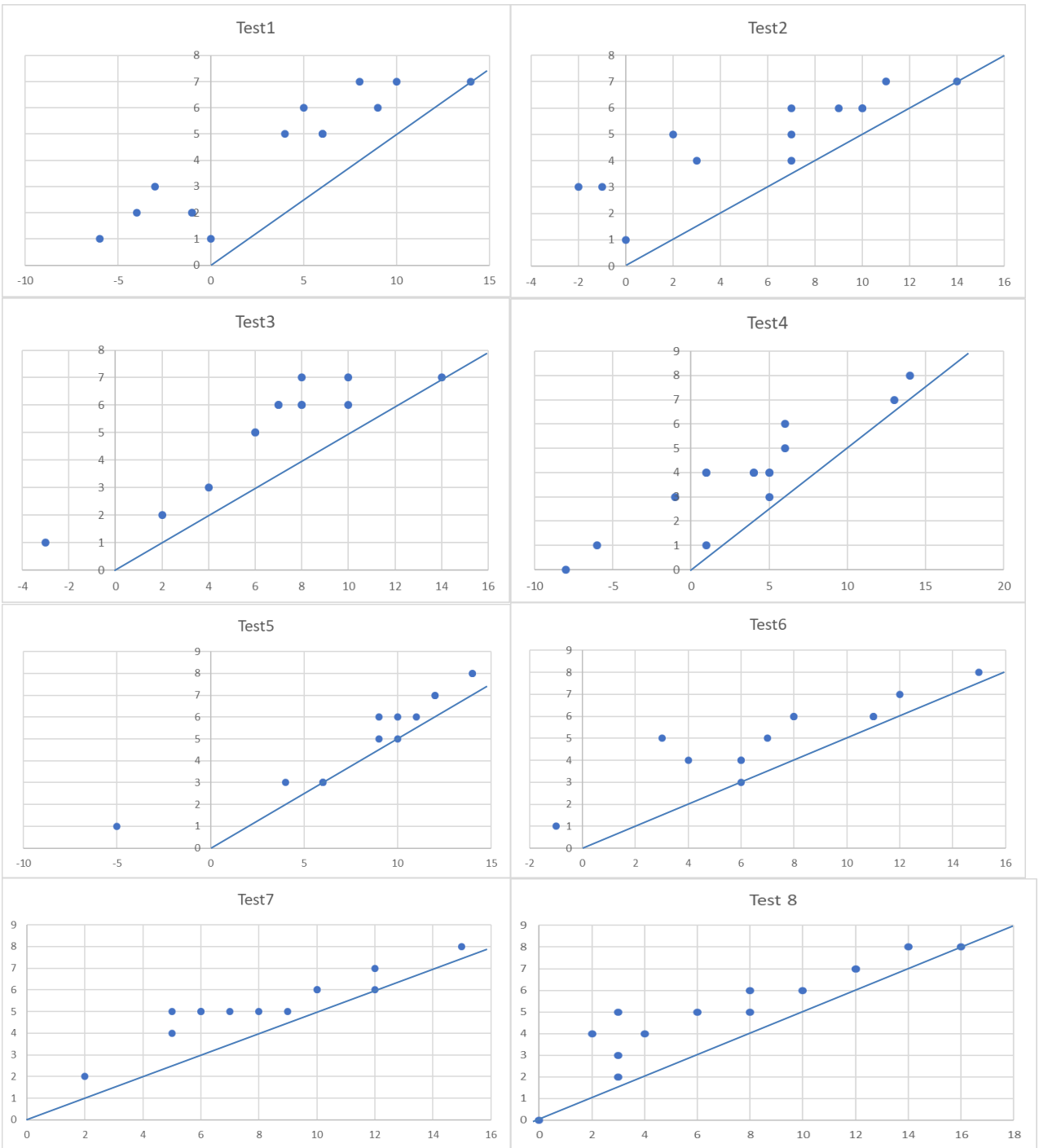

Figure 1: Test results of the CBM class (Each point represents the results of one pupil. The x-axis is the final scores, while the y-axis is the number of correct answers.)

References

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5–31.

Brensing, M., Dannewald, T., Kanzinger, A., Mayer, U., & Zender, J. (2021). Counselling in the introductory phase of studies. *Zeitschrift für Hochschulentwicklung, 16*(1), 117–136.

Brookhart, S. M. (2010). *Formative assessment strategies for every classroom: An ASCD action tool.* Association for Supervision and Curriculum Development.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In Andrade, H., Cizek, G. J. (Eds.), *Handbook of formative assessment* (pp. 3–17). Routledge. https://doi.org/10.4324/9780203874851

Foster, C. (2016). Confidence and competence with mathematical procedures. *Educational Studies in Mathematics, 91*(2), 271–288. https://doi.org/10.1007/s10649-015-9660-9

Gardner-Medwin, A. R. (1995). Confidence assessment in the teaching of basic science. *ALT-J, 3*(1), 80-85.

Gardner-Medwin, T. (2019). Certainty-based marking. In Clegg, K. (Ed.), *Innovative Assessment in Higher Education: A Handbook for Academic Practitioners*, 141–150.

Geisen, M. & Zender, J. (2023). Formative Assessment in online courses – Ideas and Experiences. In M. Ludwig, S. Barlovits, A. Caldeira, & A. Moura (Eds.), Research On STEM Education in the Digital Age. *Proceedings of ROSEDA 2023: Research Online STEM Education in the Digital Age* (pp. 25-32). WTM-Verlag.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & education, 57*(4), 2333–2351.

Hendriks, W. J. A. J., Bakker, N., Pluk, H., de Brouwer, A., Wieringa, B., Cambi, A., Zegers, M., Wansink, D. G., Leunissen, R., & Klaren, P. H. M. (2019). Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. *BMC Med Educ* 19, 178. https://doi.org/10.1186/s12909-019-1610-2

Hußmann, S., & Prediger, S. (2007). Mit Unterschieden rechnen - Differenzieren und Individualisieren. *Praxis der Mathematik in der Schule, 49*(17), 2–8.

Kanzinger, A., & Gehrig, E. (2022). Certainty-based-marking - eine kompetenzorientierte Prüfungsmethode. In IDMI-Primar Goethe-Universität Frankfurt (Eds.), *Beiträge zum Mathematikunterricht 2022.* 56. Jahrestagung der Didaktik der Mathematik. WTM-Verlag für wissenschaftliche Texte und Medien. http://dx.doi.org/10.17877/DE290R-23377

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology, 77*(6), 1121–1134.

Maier, U. (2010). Formative assessment—A promising concept for improving instruction and classroom assessment. *Zeitschrift für Erziehungswissenschaft, 13*, 293–308.

Valero, G., & Cárdenas, P. (2017) Formative and Summative Assessment in Veterinary Pathology and Other Courses at a Mexican Veterinary College. *Journal of Veterinary Medical Education, 44* (2), 331–337. https://doi.org/10.3138/jvme.1015-169R

Yuen-Reed, G., & Reed, K. B. (2015). Engineering Student Self-Assessment through Confidence-Based Scoring. *Advances in Engineering Education*, *4*(4), 1–23