# Developing a framework for assessing students’ mathematical writing

Cecilia Kilhamn1, Anna Teledahl2

1NCM, Gothenburg University, 2Örebor University, Dalarna University

Assessing students’ writing in mathematics is complex, and teachers are often left with limited support to identify and categorise quality in texts that report on problem-solving. This study is part of a larger project in which teachers and students work on discussing and improving the quality of students’ mathematical writing. For this paper, we have developed a framework for assessing the quality of students’ texts by operationalising three aspects of quality. One hundred and fifty-four student texts were analysed to develop definitions of quality levels and validate the framework. Findings show that clarity, efficiency, and use of mathematical notation in student texts can be scored independently for the problem description in students’ problem-solving texts, that differences in quality can be discerned, and that a progression in quality over time becomes visible. Some challenges are discussed.

## Introduction

Assessing the quality of students’ written mathematical work is difficult for mathematics teachers. There are several reasons for this (Casa et al., 2016; Powell et al., 2017). Firstly, it is worth noticing that assessment of students’ written documentation of problem-solving is seldom focused on the quality of their writing, separate from the quality or effectiveness of their (choice of) problem-solving strategies, making it difficult to provide students with feedback focusing exclusively on their writing skills. Secondly, there are very few examples of support for teachers in the form of research findings or curricular standards that describe aspects of mathematical writing in ways that can be used to assess different levels of quality in students’ written work. During the last two decades, research in mathematics education has paid much attention to communication, primarily highlighting oral communication, i.e., organising opportunities for students to communicate with, in, and about mathematics in group collaborations and/or whole class discussions. Research on students’ written communication appears to have taken a back seat.

In a project focused on students’ written communication, we set out to investigate students’ writing by developing a teaching design that allows teachers and students to analyse and discuss issues of quality (Teledahl et al., 2023). From these discussions and informed by previous research, we have constructed a framework in which we suggest four elements that should feature in students’ written documentation of problem solving and three aspects of quality for these elements, visible in figure 1 and described in more detail in following sections.



Figure 1. A framework for discussing quality in student written documentation.

This paper reports on the operationalisation of the framework as a tool for assessing the first element across the three aspects of quality. In a sense, it is about defining quality in student-written texts to develop a framework that researchers and teachers can use to assess quality and see progression. Our research question is thus: How can quality in the three aspects of the framework be described so that they can be used to assess problem descriptions in grade 5 student texts?

## Background

We suggested initially that assessing the quality of students’ mathematical writing is complex and that one of the reasons was that there is little support from standards or research literature. Two of the most important reasons for this are that a) writing in mathematics combines several different communicative resources to create meaning (O'Halloran, 2005), and b) what is considered good writing depends both on the local context and on conventions adopted by the global community (Barwell, 2018).

Since conventions regarding writing in mathematics are mainly concerned with mathematical notation, research has often focused almost exclusively on students’ use of mathematical symbols (Hughes et al., 2019; Teledahl et al., Forthcoming). In school mathematics, and especially in written reports of problem solving, students need to create a narrative that lets a reader understand what the problem is, what the premises are, what calculations and other procedures are carried out and what the arguments for these are as well as what the conclusion is. Students cannot create such a narrative using only mathematical notation. Instead, as O’Halloren and others have described, they are required to combine communicative resources such as natural language and visual representations with mathematical notation in ways that clearly and efficiently convey necessary information and create a comprehensive narrative (O'Halloran, 2005; Namkung et al., 2020; Powell et al., 2017; Steenrod et al., 1973).

### The framework

The framework (figure 1) is informed by previous research findings on writing in mathematics (see, for example, King et al., 2016; Kosko & Zimmerman, 2019) together with research on communicative competence in general (Rickheit et al., 2008) and several pilot studies in the initial phase of the project. The original descriptions of the four elements are as follows:

1. The *problem description* includes a description of the conditions consisting of two parts: i) premises and facts, i.e., the mathematically relevant information given in the problem and assumptions that need to be made, and ii) what is asked for in the problem, the question to be answered.
2. C*alculations* include all mathematical procedures needed to solve the problem, such as structuring, representing, modelling, calculating, or solving equations.
3. *Arguments and justifications* include making explicit the reasons behind the different steps in the solution, describing *why* specific calculations are made and where the included numbers and variables come from, as well as the logical steps that lead to a conclusion, applying logical reasoning and using implications.
4. A *conclusion* is the final response that relates to the original problem.

We expected each of these elements to appear in qualitatively different forms. Based on a literature review (Teledahl et al., Forthcoming), we initially identified four aspects of quality: clarity, efficiency, use of mathematical notation, and appropriateness. *Clarity* is related to readability and comprehension. *Efficiency* is described as communication that achieves its goal with as little effort as possible. *Mathematical notation* is often seen as the ultimate example of efficiency, but to be understood, it needs to be well-chosen and correctly used following traditional conventions. The element *appropriateness,* defined as the choice of linguistic or other resources best suited to the context and the expected reader, was later removed and is here incorporated in each of the three remaining aspects (Teledahl, 2023; Teledahl et al., 2023). Appropriateness alludes to ‘speaking the language of the classroom’ and making the different elements of the text understandable. Thus, it overlaps with the ideas of both clarity and efficiency. It is also a feature of mathematical notation, as we expect students to use mathematical notation appropriate for their grade level. Consequently, in this paper, we have dealt with issues of appropriateness in terms of what expectations we have at the specific grade level in relation to the first three aspects rather than analysing it separately as one of four aspects of quality.

### The teaching design

The data for this study is collected from a research project in which we developed a teaching design that would ensure a separation between the two processes of (i) solving a mathematical problem and (ii) communicating a solution to the problem in writing. Such a separation creates an opportunity for teachers and students to have discussions that focus exclusively on the quality of students’ writing.

The first part of our design is a problem-solving lesson, enacted by each teacher in a way the students are used to, but ending with a whole-class session in which students present and discuss different solution strategies so that, in the end, most of the students will have grasped the solution and at least one strategy for solving the problem. They are then required to produce written documentation of their problem-solving, but instead of asking students to “describe their thinking”, teachers tell them to ‘describe and justify any one of the ways the problem can be solved so that someone who has not solved the problem can understand the solution’. These documentations, henceforth called *student texts*, are collected by the teacher.

The second part of the design, in most cases as a second lesson separated in time from the first, is aimed at discussing the quality of the student texts, i.e. the focus is exclusively on the communicational quality, not the choice of problem-solving strategy. During this lesson, teachers and students analyse some of the students’ texts and discuss their merits. In the organisation of the discussions, the teachers have access to the framework described in figure 1 (initially with appropriateness added as a fourth aspect as related above). Teachers can choose to emphasise different aspects during the discussion depending on the texts produced in each classroom. After the quality discussion, students produce another text documenting the same problem, hopefully a better version than their first try.

## Method

Data collected for this study consist of student texts from five different grade 5 classes in a Swedish municipality (classroom A, B, C, D, E). As participants in the project, the teachers and students went through the teaching design working with different problems six or seven times during one semester. Each teacher collected student texts at two points in time for each problem – at the end of the problem-solving lesson and again after the quality discussion. Mostly the texts were produced as group or pair work. In addition, the teacher in classroom E let her students solve the first problem again at the end of term, collecting individual student texts at a third data point for that problem. In total, we analysed 135 student texts on the first problem called the Animal problem: 41 are from the first data point, 72 from the second, and 22 from the third data point in classroom E. In addition, the validation phase includes another 19 student texts on two other problems from three different classes.

**The Animal Problem.**

In a village there are four different kinds of animals: pigs, sheep, hens, and cows. Every fourth animal is a pig. One out of eight animals is a sheep. Half of all the animals are hens. The rest, 50 animals, are cows. How many animals of each kind are there in the village?

## Method of analysis.

To answer our research question, the framework had to be operationalised so that different levels of qualities of each aspect could be identified. We started to work with the first element, *problem description*, with the initial idea that generic qualities could be described and later used for the other elements. This initial work is reported here.

In a previous study, Teledahl (2023) used multimodal discourse analysis (Jewitt, 2011) to analyse written documentations from grade 6 and grade 11. She looked specifically at how different semiotic resources were used to create clarity and efficiency in these texts. Based on her initial findings, the work of operationalising the framework was conducted through a bottom-up approach, taking one of the three quality aspects at a time. Qualities that appeared as similar in different student texts were thematically collected into categories in an iterative process as described by Gläser-Zikuda et al. (2020). We also made note of the communicative resources used in the written documentations, distinguishing between *linguistic* resources, such as complete sentences or linking words; *visual* resources, such as headings, boundaries, arrows, bullet points, or sequencing; *mathematical symbols*; and *graphical* resources such as charts, tables, or models. Since the goal was to be able to use the framework to assess quality and progression, the categories were hierarchically organised.

Once initial categories had been identified, the whole data set for the Animal problem was assessed using these categories as levels of quality. When problems appeared, adjustments to the category descriptions were made to avoid overlap and ambiguity, after which the whole data set was re-assessed. One issue discussed was to what extent the categories were to be generically described versus specific to the problem or the element. For the framework to be useful, we expected some progression to be visible. After three iterations, final levels of quality were formulated for the element *problem description*, and assessments of the student texts were compiled. Finally, a validation of the resulting levels of quality followed, where 19 students’ texts documenting other problem solutions were assessed.

## Results

In this section, we first report how the final levels of quality came to be formulated for each for the three quality aspects. We then report the results from the assessment of all the texts using the framework, and separately for classroom E to show long-term progression. Finally, we describe the results of the validation phase.

### Quality in problem description

Analysis of the problem descriptions in the students' written documentations of the Animal problem resulted in four levels of quality for each of the three aspects *clarity*, *efficiency* and *use of mathematical notation*. The levels are numbered from 0 to 3 indicating an increased quality for higher values. In addition, the choice of communicative resources was documented.

Some interpretations related to appropriateness for the specific age and problem were necessary. We expect students in grade 5 to clarify the different proportions of pigs, hen, and sheep and the fact that there are 50 cows. Differentiating between proportion and quantity is seen as essential. To be solvable, the problem description ideally also includes the fact that there are *only* four different kinds of animals, or that the 50 cows constitute *the rest* of the animals, but we do not expect grade 5 students to discern this detail as essential to declare in the description.

#### Levels of quality for Clarity

1. No introduction of the problem (premises can be implicit in the graph or calculations).
2. Partially correct and comprehendible conditions given, or all premises given but no question, or ambiguous descriptions.
3. Conditions are given in various places of the text or *explicitly* integrated with calculations. All essential conditions are present.
4. Separation, with all essential conditions given before the calculations.

Nearly all student texts include some kind of graphical representation of the Animal problem and sometimes it is difficult to separate the problem description from the solution process. Making a graphic representation is generally interpreted as the first calculation in the solution process. But if there is no other problem description, it is assessed as the problem description with clarity level 2 if the facts are clearly and explicitly stated as premises in the graph and as level 0 if they are not.

#### Levels of quality for Efficiency

1. Difficult to make out what the conditions are (not enough information to evaluate efficiency, e.g., missing, messy, ambiguous, or incorrect premises).
2. Description of conditions including redundant text or information, for example, copying the whole problem verbatim as given.
3. Efficient presentation of essential conditions using mainly natural language and visual resources.
4. Efficient presentation of essential conditions using appropriate mathematical, symbolic, or graphical resources.

It is worth noticing that a text can be assessed as efficient on level 3 even if the mathematical notation is not entirely correct. If, for example, the equal sign is misused but the intention is understandable at the grade level in question, it is assessed as efficient. Likewise, correct spelling and grammar are not essential for assessment on level 2.

#### Levels of quality for Mathematical notation

1. No mathematical notation is used except what was given in the original problem formulation.
2. Incorrect use of mathematical notation, or use of a graphical representation that does not represent the problem correctly. (The most common misuse found in our pilot studies was an incorrect use of the equal sign.)
3. Irrelevant or redundant mathematical notation used essentially correct.
4. Relevant mathematical notation used correctly.

Many students use fractions to represent the number facts given in the problem. Here, the difference between proportions and quantities comes into play, as well as the use of the equal sign. While still accepting some mathematical notation as appropriate to grade 5 students, some nuances are assessed as incorrect. The symbol ¼ can be read as “one fourth” or “one out of four”. Ideally, it should be stated that it is “one fourth of the total number of animals”. In most cases this is accepted as taken for granted unless there is cause for misunderstanding. Some examples of mathematical notation assessed as correct are: “¼ is pigs and 50 units are cows” (In Sweden, the word “stycken” is used to specify discrete units of something); “¼ is pigs and 50 are cows”; “Pigs: ¼ … The rest, Cows: 50”. Notations assessed as incorrect are, for example: “¼ = pig, 50 = cows” (misuse of the equal sign) or “pig: ¼ … cows: 50” (ambiguous since the difference between proportion and quantity is not made clear).

### Assessment of student texts

Results of the analysis of students' problem descriptions are presented first for the 72 + 41 texts describing solutions to the Animal problem from the participating five classrooms (figure 2), then separately for all the texts from classroom E to show progression (figure 3).



a) before discussion (72 texts) b) after discussion (41 texts)

Figure 2. Distribution of texts according to assessed levels of quality across the two aspects, *clarity* and *efficiency* for the Animal problem, a) before and b) after the quality discussion. The darker the colour, the higher the frequency.

The results indicate that the framework was useful for assessing student mathematical writing quality. We can see that the students did embrace more efficient ways of communicating as a result of the teaching design focusing on quality aspects, and that their texts improved in terms of clarity. In the first diagram, many texts, are found on the lower left side on efficiency level 0 and clarity level 0 or 1. In the second diagram, most texts are assessed on level 3 for both aspects.

#### Assessing progression

We took a closer look at classroom E, where the students were given an opportunity to solve the Animal problem and report their solution in writing at the end of term. In total, 41 texts came from classroom E: seven after the first lesson (group work), twelve after the quality discussion (pair work), and twenty-two individual texts at the end of term. Figure 3 shows the distribution of texts across levels of quality for the two aspects, clarity and efficiency. In the seven texts collected after the first problem-solving lesson (figure 3a), mainly linguistic resources are used with few instances of mathematical notation. Only one text clarifies all the given conditions before the calculations and conclusion. After the quality discussion (figure 3b), almost every text clarifies the conditions using visual resources and mathematical notation, raising the level of quality for efficiency to level 3. Eleven out of twelve texts communicate efficiently using mathematical notation, albeit not always using the notation correctly.

After working with another six problems using the same teaching design, the students were again given the Animal problem to solve and 22 individually written documentations were collected (figure 3c). The predominant resources in these texts are linguistic and visual (efficiency level 2). Only seven texts include appropriate mathematical resources, suggesting that over time, awareness of clarity has increased but that all students, individually, are still not fluent with the appropriate mathematical symbols and graphical resources.



a) before discussion b) after discussion c) individually, end of term

Figure 3. Distribution of texts from classroom E according to assessed levels of quality across the two aspects *clarity* and *efficiency* for the Animal problem, a) before and b) after the quality discussion, and c) at the end of term. The darker the colour, the higher the frequency.

### Validation

To validate the operationalisation of the framework, problem descriptions of two other problems in 19 student texts were assessed by both researchers separately. A comparison showed total agreement in the assessment of quality when using the framework. However, two reflections were made, both related to mathematical notation.

In one problem, all students assigned letters to variables, for example, letting *g* and *s* represent the weight of gold and silver medals to generate the equation *g+g+s=340 gram*. In most cases, there was no explicit assignment of these letters, and thus it was interpreted as a case of incorrect use of mathematical notation assessed as level 1, while the efficiency aspect was assessed as level 3 since appropriate mathematical symbolic resources were used efficiently. Clarity was also assessed as level 3 with the argument that the essential premises concerned the relationships between the variables, not necessarily what they represented. However, one student introduced an ambiguity in the description by writing *g+g+s is 340g*, with two different meanings for the letter *g* in the equation, thus assessed as clarity level 1.

In another problem, no mathematical notation except for the numbers given in the original problem were deemed appropriate to describe the problem clearly and efficiently, therefore assessed as level 0 for use of mathematical notation. This result implies that efficiency level 3 cannot always be reached for the problem description.

## Discussion

Results from our project have shown that the described teaching design does, in fact, offer opportunities to focus exclusively on, and thereby develop, students’ ability to communicate in writing. We have also seen that access to the framework offers opportunities to disassemble writing and discuss different aspects of quality, something we believe has been missing from teachers’ practices.

The findings reported here seem to confirm that all three aspects of quality: *clarity*, *efficiency*, and *use of* *mathematical notation*, can be operationalised to enable assessment of students’ written documentations, at least concerning the problem description element. As suggested by Steenrod et al. (1973), a combination of communicative resources was valued as a quality of efficiency, and increasingly efficient when appropriate mathematical notation was included. In line with previous research, we could also see that incorrect use of mathematical notation could cause ambiguity, implying a need to highlight both how and when to use new mathematical symbols and representations. Furthermore, using the described levels of quality to assess student texts made it possible to identify differences and see progression over time. One implication of the validation phase is that all levels of all three aspects of the framework are not always applicable to every element. Correct use of mathematical notation is probably, in many cases, a more prominent aspect of the element *calculations* *and arguments* than of the problem description.

We have in this paper only developed the framework for the assessment of the first element, the problem description, enabling only an analytic view of that particulate element. When the aim is to create a comprehensive narrative, as Powell et al. (2017) suggested, the problem description needs to be separated from the solution, which is why a separation is valued higher. However, further research is needed to operationalise the framework for all elements to make a more holistic assessment possible.

Defining the levels described here turned out to be more challenging than we first expected, partly because the aspects efficiency and clarity interact with each other. One intriguing result was that we did not find any text assessed for clarity level 2 that was more than efficiency level 0. However, we did assess several texts on clarity level 1 as efficiency level 1, 2 or even 3. This can be explained by the fact that clarity level 1 is defined as being *partial*, and the part that is there could well be efficiently communicated. While premises integrated with the calculations might be efficiently communicated from a holistic point of view, it is rarely the case when assessing only the problem description.

The levels of quality described for the first element were not fully generic. However, they could serve as a useful starting point when applied to other elements, which will be the next step in our research to develop the framework further.

### References

Barwell, R. (2018). Writing in mathematics classrooms. In A. Bailey, C. Maher, & L. Wilkinson (Eds.), *Language, Literacy, and Learning in the STEM Disciplines* (pp. 1001–1114). Routledge.

Casa, T. M., Firmender, J. M., Cahill, J., Cardetti, F., Choppin, J. M., Cohen, J., & Zawodniak, R. (2016). *Types of and purposes for elementary mathematical writing: Task force recommendations*.

Gläser-Zikuda, M., Hagenauer, G., & Stephan, M. (2020). The Potential of Qualitative Content Analysis for Empirical Educational Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 21*(1).

Hughes, E. M., Markelz, A. M., & Cozad, L. E. (2019). Evaluating Various Undergraduate Perspectives of Elementary-Level Mathematical Writing. *International Journal of Science and Mathematics Education, 17*(5), 1031–1048.

Jewitt, C. (2011). Different approaches to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 28–39). Routledge.

King, B., Raposo, D., & Gimenez, M. (2016). Promoting Student Buy-In: Using Writing to Develop Mathematical Understanding. *Georgia Educational Researcher, 13*(2), 33–52.

Kosko, K. W., & Zimmerman, B. S. (2019). Emergence of Argument in Children's Mathematical Writing. *Journal of Early Childhood Literacy, 19*(1), 82–106.

Namkung, J. M., Hebert, M., Powell, S. R., Hoins, M., Bricko, N., & Torchia, M. (2020). Comparing and Validating Four Methods for Scoring Mathematics Writing. *Reading & Writing Quarterly, 36*(2), 157–175.

O'Halloran, K. L. (2005). *Mathematical discourse: Language, symbolism and visual images*. Continuum.

Powell, S. R., Hebert, M. A., Cohen, J. A., Casa, T. M., & Firmender, J. M. (2017). A Synthesis of Mathematics Writing: Assessments, Interventions, and Surveys. *Journal of Writing Research, 8*(3), 493–530.

Rickheit, G., Strohner, H., & Vorwerg, C. (2008). The concept of communicative competence. In G. Rickheit & H. Strohner (Eds.) *Handbook of communication competence* (pp. 15–62). Mouton de Gruyter.

Steenrod, N. E., Halmos, P. R., Schiffer, M. M., & Dieudonné, J. A. (1973). *How to write mathematics*. The American Mathematical Society.

Teledahl, A. (2023). Students' strategies for presenting the problem. In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)* (pp. 1695–1702). Alfréd Rényi Institute of Mathematics and ERME.

Teledahl, A., Kilhamn, C., Ahl, L., & Helenius, O. (Forthcoming). *Defining and measuring quality in students’ mathematical writing: A systematic literature review*

Teledahl, A., Kilhamn, C., Helenius, O., & Ahl, L. (2023). Writing down the solution separate from solving the problem. *For the Learning of Mathematics, 43*(2), 41–46.

### Acknowledgements

This paper is part of grant 2020-00066 under the Swedish Institute for Educational Research.