# Compromises between required and preferred features of mathematical definitions in mathematics education

Tore Forbregd, Eivind Kaspersen, Hermund André Torkildsen and Trygve Solstad

Norwegian University of Science and Technology, Trondheim, Norway; tore.a.forbregd@ntnu.no, eivind.kaspersen@ntnu.no, hermund.a.torkildsen@ntnu.no, trygve.solstad@ntnu.no

*The mathematics education literature distinguishes between the preferred and the required features of mathematical definitions. However, it is unclear whether and how this distinction differs between mathematical and didactical contexts. Here, we address the didactical context in which the purpose of the definition is to help students learn and understand the meaning of a new concept. We administered a comparative judgement study to assess how 12 mathematics teacher educators value the relative importance of required and preferred features of mathematical definitions. We found that the educators did not value the required features as more important than the preferred features of mathematical definitions. Furthermore, we found that the educators valued some required features more than others. These results suggest that, in a didactical context, the categorization into preferred and required features of mathematical definitions may not be as clear cut as indicated in the mathematics education literature and depends on the context or purpose of the definition.*

*Keywords: definitions, comparative judgement, teacher education, teacher educators, teacher lecturers*

## Introduction

To develop proficiency in mathematical reasoning, students must understand the essential ingredients of mathematics; one of these ingredients is *mathematical definitions*. Research shows, however, that students struggle with definitions (Edwards & Ward, 2004), partly because definitions are multi-faceted by nature. Morgan (2006), for instance, explained that definitions are used both for *deductive reasoning* and to *understand* a new concept. In previous work we found that scholars in mathematics education commonly label some features of mathematical definitions as *required* and others as *preferred* but not required (Forbregd et al., Submitted).

However, students and teachers of mathematics must balance these, sometimes, contrasting features of mathematical definitions. The literature does not address how the categorization into preferred and required features may depend on the purpose of the definition, such as whether it is intended for logical reasoning in a pure mathematics context or for learning and understanding a new concept in a didactical context. As a first step to clarify this issue, this paper addresses the didactical context by assessing how university lecturers who train mathematics teachers value the required and the preferred features of mathematical definitions.

A naïve interpretation of the categorization is that (1) if all preferred features can be omitted from a definition, every required feature is more important than every preferred feature, and (2) if no required feature can be omitted from a definition, all required features are equally important. Using comparative judgement, we answer these corresponding research questions:

1. *How do Norwegian mathematics teacher lecturers balance the required and the preferred aspects of mathematical definitions?*

2. *How do Norwegian mathematics teacher lecturers value the relative importance of the required aspects of mathematical definitions?*

We point out that this is a preliminary study which is part of a larger project.

## Aspects of mathematical definitions

From a literature review (Forbregd et al., Submitted), we found five main themes on how scholars in mathematics education ($N = 74$) describe mathematical definitions. In this study, we focus on two of these themes, namely, the required and the preferred aspects of mathematical definitions.

In the review, we found three strands of the required features of mathematical definitions. First, several scholars emphasise *formal* requirements, for instance, that definitions must be consistent and non-contradicting (e.g., Johnson et al., 2014) and that they must be unambiguous (e.g., Foster & de Villiers, 2016). Second, multiple authors describe *existence* as a required aspect of definitions. That is, for every definition, at least one example or instance must exist (e.g., Avcu, 2019). Also, although definitions exist within representation systems, the definitions should be invariant under change of representations (e.g., Sánchez & García, 2014). Third, all mathematical definitions must exist within deductive systems (e.g., Cansiz Aktaş, 2016). Hence, all mathematical definitions employ only previously defined concepts (Van Dormolen & Zaslavsky, 2003).

The review identified three strands also for the preferred features. Definitions should be minimal, a feature that was highlighted by, for example, Vinner (1991). In our study, we coded *minimality* as a preferred feature, although we appreciate that scholars disagree on this issue (e.g., Van Dormolen & Zaslavsky, 2003). A second preferred strand comprises *aesthetic* features, such as elegance (e.g., Zazkis & Leikin, 2008), precision (Levenson, 2012), and clarity (Leikin & Winicki-Landman, 2000). Finally, definitions should be didactically suitable (Winicki-Landman & Leikin, 2000), intuitive (Ouvrier-Buffet, 2011), and match students' knowledge and needs (Leikin & Winicki-Landman, 2000).

## Methods

From the aspects discussed above, we gathered 31 statements—each corresponding to either a required or a preferred feature of mathematical definitions. The statements were quotes from the literature about mathematical definitions; however, to reduce bias, we removed words such as ”must” and ”should” from the statements. The statements were also rephrased so that they had roughly the same wording. Examples of statements are: ”That it is consistent and non-contradicting” and ”That it matches the target group's knowledge and needs”. See Table 1 for a complete list of statements.

The statements were uploaded to the web application ”No More Marking” (abbr. NMM) (Wheadon, 2020). Respondents had to meet the criteria of being active teachers in mathematics teacher education at university level. Subsequently, 12 assistant professors in mathematics teacher education were enlisted as judges to conduct a comparative judgement study.

Each judge conducted 40 comparisons[1] of two statements—one pair at a time—with the following question: ”Which of the two statements about mathematical definitions are more important in mathematics education?”. For each judging session the pairings of statements to be compared were

[1] There were no possibility for ranking statements as equally important, that is a judge must rank one statement as more important than the other.

randomized by NMM[2]. The data were collected on Android tablets. The tablets were administered by the researchers, and hence, no person-identifying data or IP addresses were stored. The respondents compared 480 pairs of statements in total.

In the analysis, we used a Rasch model for paired comparison (e.g., Wright & Stone, 1979). This model expresses the likelihood that a statement $n$ with measure $\beta_n$ beats a statement $m$ with measure $\beta_m$ as

$$P(n \text{ beats } m) = \exp(\beta_n - \beta_m) / [1 + \exp(\beta_n - \beta_m)] \quad (1)$$

The numerator (Eq. 1) informs that, as $\beta_n$ increases relative to $\beta_m$, the likelihood that $n$ beats $m$ increases. The denominator ensures that the likelihood is constrained between zero and one. The comparison data were used to fit the Rach model, yielding a (logit) score for each statement.

To assess the level of agreement between the judges, we conducted two analyses: First, we assessed the loss of invariance from differential item functioning (DIF). Specifically, we allocated each judge into one of two groups, and then we assessed whether the reported measures depended on which group we used in the analysis. Here, we used the Rasch-Welch $t$-test with a critical $p$-value of .05. Second, we examined the judge Infit Mnsq. Infit Mnsq is based on mean square standardised residuals of observations and the Rasch model (Bond et al., 2020). Roughly, persons with Infit Mnsq higher than one tend to respond unpredictably relative to the rest of the respondents. The analysis of Infit Mnsq was also conducted for each statement, and here, Infit Mnsq greater than one indicates that the sample of respondents disagreed on the relative importance of this statement.

Finally, after we had estimated and validated the measures of the statements, we conducted two tests, each responding to the research questions. To examine the null-hypothesis that required and preferred features were valued equally by the respondents (RQ1), we conducted a $t$-test on the mean values of the required and the preferred aspects. To test the null hypothesis that all required features of mathematical definitions were valued equally (RQ2), we conducted 1000 simulations in R (R Core Team, 2020) using the *Rwinsteps* package (Albano & Babcock, 2019). Here, we constrained the measures of every required statement as equal, and then, for each simulation, we returned the standard deviation (*SD)* of the required statements after 480 simulated comparisons. These simulated *SD*s were compared with the *SD* in the empirical data.

# Results

## Teacher lecturers' compromises between required and preferred features of definitions

Scholars in mathematics education distinguish between features of definitions that are required and features that are preferred but not required. As we have argued, if we applied only mathematical principles when we evaluated a definition, we should value all the required features more than the preferred features. However, the empirical results in this study show that lecturers in mathematics education balance these features. That is, it is not always the case that a required feature is regarded as more important than a preferred feature of a mathematical definition.

Examples of such compromises are shown in Table 1. As a first observation, we see that the two features that the lecturers valued highest were both labelled as required. Accordingly, the respondents

[2] The randomization algorithm aims to equalise the frequency of each statement in the total number of comparisons.

in this study were not willing to prioritise didactical principles (e.g., that the definition should be didactically suitable for the target group) at the expense of these two required features.

The following four features, however, were labelled as preferred. When the lecturers compared one of these preferred features (e.g., that the definition matches the target group's knowledge and needs) with required features of lower measures (e.g., that a definition fits into and is part of a deductive system), the teacher educators were more likely to select the former than the latter.

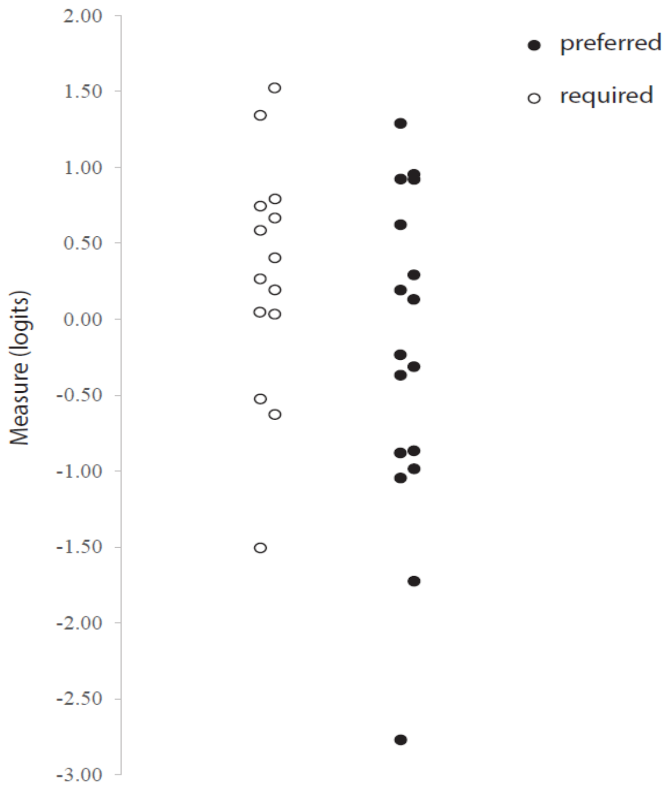


**Figure 1. Values of required and preferred features of mathematical definitions**

Moving beyond particular examples, the results, which are summarised in Figure 1, contradict the hypothesis that, in education, required features are always perceived as more important than preferred features of mathematical definitions. That is, we found no significant difference ($p$ = .14) in how lecturers in mathematics education valued the required features ($M$ = 0.28 , $S.E.$ = 0.21) and how they valued the preferred features ($M$ = -0.23 , $S.E.$ = 0.26) of mathematical definitions.

**Table 1. The relative importance of required and preferred features of mathematical definitions**

| Measure | s.e. | Infit Mnsq | Aspect |
|---|---|---|---|
| 1.52 | 0.34 | 1.0 | R That it is well-defined, that is to say, the meaning is unambiguous. |
| 1.34 | 0.32 | 1.0 | R That it is consistent and non-contradicting. |
| 1.29 | 0.33 | 1.1 | P That it matches the target group's knowledge and needs. |
| 0.95 | 0.30 | 0.9 | P That it captures and synthesises the mathematical essence of the concept. |
| 0.92 | 0.30 | 0.9 | P That it is didactically suitable to the target group. |
| 0.92 | 0.30 | 0.9 | P That it is understandable to the target group. |
| 0.79 | 0.29 | 0.9 | R That if there are multiple definitions for a given concept, they must be mathematically equivalent. |
| 0.74 | 0.30 | 1.0 | R That it allows to discriminate between instances and non-instances. |
| 0.66 | 0.29 | 1.1 | R That it is consistent with the mathematical theory formed thus far. |
| 0.62 | 0.29 | 0.9 | P That it only mentions necessary terms and properties so that it is possible to distinguish an instance from a non-instance. |
| 0.58 | 0.29 | 1.1 | R That it has a unique interpretation; in other words, it is well-defined and unambiguous. |
| 0.40 | 0.29 | 0.9 | R That it can be proven that at least one instance of the defined concept exists. |
| 0.29 | 0.29 | 1.1 | P That it has easily identifiable examples. |
| 0.26 | 0.29 | 1.1 | R That it is part of a deductive system, in a hierarchical and noncircular manner within itself and across existing axioms, definitions, and theorems. |
| 0.19 | 0.29 | 1.0 | P That it is hierarchical, in the sense that the terms used in the definition are known to the target group. |
| 0.19 | 0.28 | 1.1 | R That it allows instances and non-instances of the concept to be discriminated with certainty, consistency, and efficiency. |
| 0.13 | 0.29 | 1.1 | P That it only employs previously defined concepts known to the target group. |
| 0.04 | 0.28 | 1.0 | R That at least one example of the defined concept exists. |
| 0.03 | 0.28 | 0.9 | R That all the properties stated in the definition can coexist. |
| −0.24 | 0.29 | 1.0 | P That it is precise. |
| −0.32 | 0.29 | 1.0 | P That it is clear. |
| −0.37 | 0.29 | 1.0 | P That it does not contain properties which can be mathematically inferred from other parts of the definition, i.e., it is minimal. |
| −0.53 | 0.30 | 0.8 | R That equivalence can be proven if more than one definition is given for the same concept. |
| −0.63 | 0.30 | 1.1 | R That it fits into and is part of a deductive system. |
| −0.87 | 0.31 | 1.1 | P That it is useful, for example, for proving theorems. |
| −0.88 | 0.31 | 0.9 | P That it is minimal. Minimality means that no conditions in a definition can be inferred from the other conditions; that is, there is no redundancy. |
| −0.99 | 0.32 | 0.9 | P That it does not contain superfluous words or symbols, and that it "looks nice". |
| −1.05 | 0.32 | 1.2 | P That it is intuitive. |
| −1.51 | 0.35 | 1.1 | R That it describes any new concept as a special case of a more general concept. |
| −1.73 | 0.37 | 1.1 | P That the name of the concept must be closely related to its natural-language usage. |
| −2.77 | 0.53 | 0.8 | P That it is elegant. |

Moreover, the results, presented in Table 1 and Figure 1, indicate that, in education, some required features are valued more strongly than other required features. To see if this difference was statistically significant, we compared the empirical standard deviation (*SD)* of the required features (*SD* = 0.90) to a simulated dataset in which all required features were modelled as equally important. Under simulated conditions, where all *true* measures are equal (i.e., when the *true SD* is zero) and all respondents make their judgements entirely in accordance with the Rasch model, we found that the empirical *SD* that can be expected due to measurement errors was 0.61. Since this difference was significantly different from the SD of the empirical measure ($p$ =.04) we conclude that the reported *SD* in the required features is not due to measurement errors alone.

**On the stability of the results**

The judge Infit Mnsq values (not to be confused with *item* Infit Mnsq in Table 1) indicate the extent to which the lecturers agreed about the relative importance of the aspects of mathematical definitions. In our study, most Infit Mnsq values were around one, which is the expected value when all judges value the aspects equally. One judge stood out, however, with an Infit Mnsq value of 1.5. Although this value suggests that the respondent valued some of the aspects differently than most of the other respondents, the difference did not have practical consequences for the results on which we report in this paper. The correlation between the measures in the full sample and the measures when this respondent's judgements were excluded was $r$ = .99.

To assess the stability of the results in more detail, we split the sample of respondents randomly into two groups. A DIF analysis on these groups showed that two features had measures that differed significantly ($p$ = .03) between the groups. These features were ”That at least one example of the defined concept exists” and ”That it is useful, for example, it is useful for proving theorems”. Apart from these aspects, the respondents seemed to have an overall agreement on the relative importance of the features of mathematical definitions.

The reliability (analogous to Cronbach's alpha) of the measures was .77. Roughly, this value suggests that, while we obtain measures for individual features of mathematical definitions, we cannot make fine-grained inferences. For instance, within a cluster of features around zero logits, we cannot tell whether the respondents valued some of the features more highly than others.

## Discussion

There are many features to consider when one formulates, or chooses, a definition formulation for a given mathematical concept in an educational setting. This study gives a glimpse into the ranking of features of mathematical definitions that teacher educators must take into account when teaching mathematics.

That some features of mathematical definitions are labelled as *required* in the mathematics education literature, can give the impression that there is no leeway for lecturers in mathematics education to emphasise preferred features of mathematical definitions at the expense of required features. However, we have shown that Norwegian lecturers in mathematics education valued many preferred features—especially those of a didactical flavour—as more important than mathematical requirements (e.g., that a mathematical definition must be part of a deductive system). Furthermore, we have shown that mathematical requirements are not necessarily valued equally by mathematics educators, suggesting that the required features are not all regarded as required when the purpose of the definition is to learn and understand new concepts.

The ranking of statements in Table 1 reveals a thematical distinction of the statements. Many of the aspects on the higher end in Table 1 relate to the clarity of the concepts with respect to the target group. By contrast, many of the statements on the lower end are non-functional features and applications of definitions (e.g., that they are ”useful”, ”elegant”, and ”special cases of more general concepts”). Another point that may be inferred from Table 1 is that statements with a clear didactical agenda are held in higher regard than those which do not (e.g., ”That it is hierarchical, in the sense that the terms used in the definition are known to the target group” was perceived as more important than ”That it describes any new concept as a special case of a more general concept”).

In a pure mathematical context, it is possible that there are no or few compromises to be made between requirements and preferred features. However, in a didactical context, we have shown that there is a measurable tension between mathematically important features and didactically important features of mathematical definitions. A ”good” definition in pure mathematics, might not be considered ”good” didactically, while a ”good” didactical definition might not be ”good”, perhaps not even valid, mathematically. Mathematics educators are willing to relax some mathematical requirements to allow for some didactical considerations. Insofar as some of these compromises are necessary, teachers and lecturers in education are faced with a challenging task: to teach mathematics by relaxing requirements of mathematical definitions.

Our work raises some new research questions. A larger study on the aspects of definitions from the literature seems worthwhile, both in an educational context as in our case and in a pure mathematical context. For example, it is not clear to us that all of the features labelled as *required* in the mathematics education literature are literally or objectively required in the purely mathematical context. Another relevant question is whether features can be usefully characterized as more ”didactical” or more ”mathematical”. Qualitative studies could further help to explain, exemplify, and explicitly describe our observations and the tension between mathematical features and didactical features of definitions. This would hopefully be valuable for teachers and educators in their work on mathematical definitions and introduction of new concepts for students to use in their deductive reasoning and proving.

## References

Albano, A. & Babcock, B. (2019). *Rwinsteps: Running Winsteps in R* (1.0-1.1) [Computer software]. https://CRAN.R-project.org/package=Rwinsteps

Avcu, R. (2019). A comparison of mathematical features of Turkish and American textbook definitions regarding special quadrilaterals. *International Journal of Mathematical Education in Science and Technology*, *50*(4), 577–602. https://doi.org/10.1080/0020739X.2018.1529338

Bond, T., Yan, Z. & Heene, M. (2020). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Routledge.

Cansiz Aktaş, M. (2016). Turkish high school students’ definitions for parallelograms: Appropriate or inappropriate? *International Journal of Mathematical Education in Science and Technology*, *47*(4), 583–596. https://doi.org/10.1080/0020739X.2015.1124931

Edwards, B. & Ward, M. B. (2004). Surprises from mathematics education research: Student (mis)use of mathematical definitions. *The American Mathematical Monthly*, *111*(5), 411–424. https://doi.org/10.1080/00029890.2004.11920092

Forbregd, T. A., Torkildsen, H. A., Kaspersen, E. & Solstad, T. (Submitted). Towards a unified account of definitions in mathematics education research: A systematic literature review.

Foster, C. & de Villiers, M. (2016). The definition of the scalar product: An analysis and critique of a classroom episode. *International Journal of Mathematical Education in Science and Technology*, *47*(5), 750–761. https://doi.org/10.1080/0020739X.2015.1117148

Johnson, H. L., Blume, G. W., Shimizu, J. K., Graysay, D. & Konnova, S. (2014). A teacher's conception of definition and use of examples when doing and teaching mathematics. *Mathematical Thinking and Learning*, *16*(4), 285–311. https://doi.org/10.1080/10986065.2014.953018

Leikin, R. & Winicki-Landman, G. (2000). On equivalent and non-equivalent definitions: Part 2. *For the Learning of Mathematics*, *20*(2), 24–29.

Levenson, E. (2012). Teachers' knowledge of the nature of definitions: The case of the zero exponent. *The Journal of Mathematical Behavior*, *31*(2), 209–219. https://doi.org/10.1016/j.jmathb.2011.12.006

Morgan, C. (2006). What is a definition for in school mathematics? In M. Bosch (Ed.), *European Research in Mathematics Education IV: Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education.* (pp. 861–871). FUNDEMI IQS – Universitat Ramon Llull. https://discovery.ucl.ac.uk/id/eprint/10000418/

Ouvrier-Buffet, C. (2011). A mathematical experience involving defining processes: In-action definitions and zero-definitions. *Educational Studies in Mathematics*, *76*(2), 165–182. https://doi.org/10.1007/s10649-010-9272-3

R Core Team. (2020). *R: A language and environment for statistical computing* (3.6.3) [Computer software]. R Foundation for statistical computing.

Sánchez, V. & García, M. (2014). Sociomathematical and mathematical norms related to definition in pre-service primary teachers' discourse. *Educational Studies in Mathematics*, *85*(2), 305–320. https://doi.org/10.1007/s10649-013-9516-0

Van Dormolen, J. & Zaslavsky, O. (2003). The many facets of a definition: The case of periodicity. *The Journal of Mathematical Behavior*, *22*(1), 91–106. https://doi.org/10.1016/S0732-3123(03)00006-3

Vinner, S. (1991). The role of definitions in teaching and learning. In D. Tall (Ed.), *Advanced Mathematical Thinking* (pp. 65–81). Kluwer academic.

Wheadon, C. (2020). *No More Marking*. https://www.nomoremarking.com/

Winicki-Landman, G. & Leikin, R. (2000). On equivalent and non-Equivalent definitions: Part 1. *For the Learning of Mathematics*, *20*(1), 17–21.

Wright, B. D. & Stone, M. H. (1979). *Best test design*. MESA Press.

Zazkis, R. & Leikin, R. (2008). Exemplifying definitions: A case of a square. *Educational Studies in Mathematics*, *69*(2), 131–148. https://doi.org/10.1007/s10649-008-9131-7