MATHEMATICS PAPERS

SHORT PAPERS

Maths004

USING A COMPARATIVE JUDGEMENT APPROACH TO ASSESS THE PROBLEM-SOLVING SKILLS OF PRIMARY SCHOOL PUPILS P Barmby¹, C Foster², I Jones², J Kelly³, J Milinkocic⁴ No More Marking¹, Loughborough University², University of Cambridge³, United Kingdom & University of Belgrade⁴, Serbia

Abstract

Comparative judgement has been put forward as a way of assessing more open responses to mathematical questions, for example in problem solving. This paper describes a small-scale study involving a comparative judgement assessment of the problem-solving skills of 17 pupils from one primary school in England, with judgments provided by 10 teachers. The Scale Separation Reliability was 0.87, showing a high degree of accuracy in differentiating between the quality of responses. Examples of responses showed a valid progression in the problem-solving skills shown. Questionnaire responses from the teacher judges support the potential of the process to inform teachers and pupils of the range of approaches that can be used in such a task. In addition, through the built-in moderation process, comparative judgement can provide formative support in aligning teachers' views of problem solving.

Keywords: Assessment, Comparative judgement, Primary, Problem Solving

Introduction

There is increasing interest in comparative judgement as an assessment method in mathematics (e.g. Marshall, Shaw, Hunter & Jones, 2020). In contrast to criteria-based approaches to assigning scores to responses, comparative judgement involves iteratively comparing pairs of responses, choosing the better response based on the experience of the judge.

We describe a small-scale study looking at the feasibility of carrying out a comparative judgement assessment of the problem-solving skills shown by primary-age pupils in England. The paper presents: (a) the results of the assessment, including its reliability, and an analysis of the characteristics of the responses; (b) the results of a questionnaire to judges examining the accuracy and usefulness of the approach.

Literature review

Problem solving is regard as a process that occurs through a number of stages (e.g. Polya, 1973; Schoenfeld, 1985). A problem is a task for which learners have no readily available procedure, and learners must develop strategies for solving a task in a systematic manner (NCTM, 2000). The identification of a problem is linked to the solver's knowledge and abilities in the moment of solving the problem (Milincović, 2015).

However, such tasks are difficult to assess (Sangster, 2012). They may have more than one answer, or more than one approach to a solution. To avoid this difficulty, it is tempting to use more closed problems, although this goes against the reasons for using open approaches to mathematics (Blum & Niss, 1991). One possible approach to assessment, allowing for the range of approaches to tackling a problem, is comparative judgement. This is where repeated pair-wise comparisons of responses by judges result in a measurement scale of the quality of these responses (Pollitt, 2012). In mathematics, Ian Jones and colleagues at Loughborough University have put forward comparative judgement as a way of looking at understanding (Bisson, Gilmore, Inglis & Jones, 2016) and problem solving (Jones & Inglis, 2015). In particular, comparative judgement is suitable for "assessing nebulous constructs that are deemed important but which are difficult to specify comprehensively in mark schemes" (Jones & Inglis, 2015, p. 341). Comparative judgement has been used successfully in primary education to assess free-response tasks in mathematics (Hunter & Jones, 2018).

Methodology

The question used for this study was based on a problem used by Elia, van den Heuvel-Panhuizen and Kolovou (2009) in their study of pupils aged 9 to 10 years in the Netherlands. It was chosen because a reasonable proportion (40%) of pupils in that study managed to answer the question correctly and would more likely be a problem for our pupil cohort without being too difficult. The wording and numbers in the question were slightly changed to avoid ambiguity and make it more difficult for pupils to reach the answer by trial and error:

Mary has 22 tokens. Each token can be worth 5 or 10. Mary's tokens are worth 150 altogether. How many tokens of each type does Mary have?

Pupils were asked to restrict their response to one side of A4 paper. This was to facilitate the subsequent upload of the image of the response via the mobile upload facility¹ of No More Marking's website. 17 Year 5 pupils (aged 9-10) from one primary school in England uploaded responses, the school being conveniently chosen as one of the teachers volunteered the school to participate. 10 teachers from the school were then asked to carry out 21 comparative judgements each between pairs of responses, choosing "the better problem-solving approach" in each case (Figure 1).



Figure 1: Comparative judgement of a pair of responses by choosing left or right

¹ The mobile upload facility was used due to uncertainty of school openings due to Covid 19.

Following the judging, each teacher was sent questions in an email questionnaire looking at the accuracy of the assessment, the judging process, and the usefulness of the process for schools and teachers. Nine out of the 10 judges provided responses to these questions. These responses were examined, common issues identified, and these were taken forward as possible emerging themes.

Findings

On completion of the judging, the Scale Separation Reliability was 0.87 (and split-half Inter-Rater Reliability of 0.81 based on 100 iterations), showing a high degree of accuracy in differentiating between the quality of the responses. Looking at the progression in the responses with increasing scaled score, five responses at intervals of 25 scaled score points (on a scale of 0 to 100) were examined (Figure 2). To exemplify this progression, three characteristics of the responses were identified by the researchers: a correct response; a systematic approach, and a clear explanation of the approach (Table 1).



Figure 2: Selection of responses at scaled score intervals (scaled scores shown)

Table 1: Progress in characteristics of responses from the researchers' impressions

Scaled Score	Correct Response?	Systematic Approach?	Clear explanation
0	×	*	×
30	\checkmark	×	×
51	\checkmark	\checkmark	×
63	\checkmark	\checkmark	\checkmark
100	\checkmark	\checkmark	\checkmark

The scoring from the comparative judgement process seems to show progress not only in whether a correct response was reached, but also in the approach and explanation from the pupils. This suggests that comparative judgement by these teachers resulted in a valid ordering the quality of problem-solving responses from these pupils.

Examining the judges' views of the comparative judgement process, no teacher disagreed with the accuracy of the process, and four judges specifically mentioned that they saw it as providing an accurate reflection of pupil abilities. Seven teachers also highlighted that the process assessed broader skills such as understanding and thinking. And in connection with this, five teachers identified the benefit of the process showing the different approaches that could be used, and subsequently teachers being able to show or model these approaches to develop pupils' problem-solving skills. In addition, the issue of the process moderating assessments across a group of teachers was highlighted by two teachers, supporting them in aligning their view of problem solving with those of others.

Discussion

The statistical reliability for the task, along with the progression shown in the responses, suggests that comparative judgement can provide an accurate and valid approach to assessing pupils' problem-solving skills (Jones & Inglis, 2015). Furthermore, due to the suitability of the approach to assessing open-ended responses, it avoids some of the constraints of assessing problem solving skills highlighted in the literature (Blum & Niss, 1991).

An interesting aspect that emerged from the views of the teacher judges was the formative potential of the process to teachers and subsequently to pupils, by highlighting the range of possible responses to a task, and also the possibility of aligning views of problem solving across a group of teachers. The comparative judgement process can therefore provide support for teachers in how they view problem solving, and this formative aspect will be explored further in our future research.

References

- Bisson, M. J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education*, 2(2), 141-164.
- Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects - State, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, 22(1), 37-68.
- Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. ZDM, 41(5), 605.
- Hunter, J., & Jones, I. (2018). Free-Response Tasks in Primary Mathematics: A Window on Students' Thinking. In Hunter, J., Perger, P., & Darragh, L. (Eds.), Making waves, opening spaces (Proceedings of the 41st annual conference of the Mathematics Education Research Group of Australasia), pp. 400-407. Auckland: MERGA.
- Jones, I. & Inglis, M. (2015). The problem of assessing problem solving: can comparative judgement help? *Educational Studies in Mathematics*, 89, 337-355.

- Marshall, N., Shaw, K., Hunter, J., & Jones, I. (2020). Assessment by comparative judgement: an application to secondary Statistics and English in New Zealand. *New Zealand Journal of Educational Studies*, 1-22.
- Milinković, J. (2015). Conceptualizing problem posing via transformations. In F. M. Singer, N. F. Ellerton, & J. Cai (Eds.), *Mathematical problem posing: From research to effective practice* (pp. 47-70). NY: Springer.
- National Council of Teachers of Mathematics (NCTM) (2000). Principles and standards for school mathematics. Reston, VA: NCTM.
- Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*(3), 281-300.
- Polya, G. (1973). How to solve it. Princeton, NJ: Princeton University Press.
- Sangster, M. (2012). The rise and fall of an investigative approach to mathematics in primary education. *Proceedings of the British Society for Research into Learning Mathematics, 32*(1).
- Schoenfeld, A. H. (1985). Mathematical problem solving. New York: Academic Press.