Towards an empirical validation of mathematics teachers' intuitive assessment practice exemplified by modelling tasks

Xenia-Rosemarie Reit

Goethe-University Frankfurt, Institute for mathematics education, Germany; reit@math.uni-frankfurt.de

The identification of intermediate steps in student solutions as a basis for assessment is a common procedure in mathematics teaching. Modelling tasks, providing more than one solution approach, are considered hard to assess. This is not least a reason for the unsatisfactory proportion of modelling in school. Assuming that task difficulty is strongly connected with assessment, the implications of a study about the difficulty of modelling tasks are discussed. Starting point for the discussion is the question whether intuitive assessment practices like the identification and scoring of intermediate steps, can be supported by empirical findings. The focus is on the influence of cognitive aspects regarding structural characteristics of solution approaches. Results indicate that a pure sequential consideration of thought structures in a solution approach lead to reasonable results and might justify its application in school due to its straightforward implementation.

Keywords: Assessment, cognitive structure, modelling tasks, mathematics teaching.

Introduction

Within the mathematics community it is mostly agreed upon the positive impact of modelling tasks on the learning of students. Mathematical modelling is promoted and there are many votes for its broader implementation in school mathematics. However, several studies provide evidence that modelling is far away from playing an integral role in everyday school teaching, in Germany and also elsewhere (Blum, 2007, p. 5). Jordan et al. (2006) confirm that the proportion of modelling in daily school routine is low. Research focusing on the teachers’ point of view reveals several difficulties that teachers are confronted with. In a study of Schmidt (2010) it has been found out that 67% of the interviewed teachers indicate assessment as being the major challenge in the implementation of modelling tasks. Blum (1996) also speaks of an increased difficulty in the context of modelling tasks. These findings are comprehensible in view of the multiple solution approaches of modelling tasks.

A common opinion is that modelling tasks cannot be assessed as objectively as traditional task formats (Spandaw & Zwaneveld, 2010). However, if we want modelling to be part of mathematics teaching, it must part of the grading (Hall, 1984). With the striking quotation “What you assess is what you get” Niss (1993) also argues in favor of a provision for modelling activities in the grading. The aim of the present paper is not to discuss and contrast formative and summative assessment since the advantages of formative assessment as assessment for learning movement could be confirmed in different settings (e.g. Black & William, 1998) and are not denied. However, in view of the fact that summative assessment aspects hinder the implementation of modelling tasks in everyday school live, it is necessary to provide tools or possibilities in that direction.

Besides a number of assessment methods which aim at assessing modelling competence (e.g. Berry & Le Masurier, 1984; Haines, Crouch, & Davis, 2000), there is hardly any assessment instrument
which can be used for assessing modelling tasks in everyday school life. In this context so far only Maaß suggests an assessment scheme which can be adapted to different modelling tasks by a variable weighting of several categories (Maaß, 2007, p. 40). However, an empirical validation is lacking such that the assessment scheme might serve as orientation but it cannot give detailed instruction.

On the way to an assessment scheme for a mathematics task, a common procedure of mathematics teachers (in the following referred to as “intuitive assessment practice”) is to identify reasonable intermediate steps in a solution which are worthwhile to be scored. On that basis an assessment scheme is set up which determines the conditions to be fulfilled for a differentiated scoring of those intermediate steps. Hence, there is a procedural difference between the phase of identifying scoreable aspects in a solution and an assessment scheme. The former is the requirement for the latter. At this point the present paper ties on by discussing the use of so called thought structures to identify reasonable intermediate steps in solution approaches of modelling tasks in connection with its difficulty. The question of identifying intermediate steps and the influence of their structure within a solution approach to its difficulty has been analysed by Reit (2016). In this study different models are developed and evaluated to determine the difficulty of solution approaches. Assuming that assessment of a mathematics task is strongly determined by its difficulty, interesting conclusions can be drawn concerning common assessment practice in school. Results of the study of Reit (2016) indicate that there is quantifiable influence of structural characteristics of a solution approach on its difficulty. However, it is also stated that a sequential model which is based upon a sequential arrangement of thought operations, similar to the intuitive assessment practice of mathematics teachers, can also be confirmed.

**Theoretical framework**

The core of the study of Reit (2016) is a structural analysis of students’ solution approaches of modelling tasks. These thought structures of solution approaches indicate the chronology of thought operations to be done to arrive at a solution. Assuming that parallel thought operations complicate a solution approach, a non-weighting difficulty model (addition model) is contrasted with four models varying in their weighting of parallel thought operations.

**Thought structure analysis**

Recalling structures is a wide-spread procedure in mathematics (Bourbaki, 1961, pp. 163). In this context Breidenbach (1963) looks at the structural-substantial complexity of a word problem to decide amongst others about its difficulty. He formulates that tasks with one operation deal in the simplest case, with one issue in which three factors play a role and every factor is uniquely determined by the two others (Breidenbach, 1963, p. 200). Breidenbach named such tasks Simplex. A linking of several Simplex is called Komplex. Further developments of Winter and Ziegler (1969) lead to the arithmetic tree which is still used in mathematics textbooks (Figure 1). An obvious but so far empirically not validated conclusion is that a larger number of Simplex and a more complicated nesting of them, has an effect on the difficulty of the tasks’ solution (Graumann, 2002).
Figure 1: Arithmetic tree of an exemplary task following Winter and Ziegler (1969)

The study of Reit (2016) investigates the cognitive complexity of a solution approach on the basis of its structural complexity represented by its arithmetic tree-like structure. At that point the coherence of structural considerations and cognitive psychological theories play an important role. In a study of Fletcher and Bloom (1988) it is assumed that text comprehension is a kind of problem solving process, where the reader must find a causal chain which links start and end of a text. Furthermore they assume that information must be kept simultaneously in the working memory to be able to form such a causal chain. Results of their study show that readers must keep that information available that is the direct predecessor in the causal chain. It can be concluded that the task of the working memory is to keep information available which is necessary to link old and new information (Baumann 2000).

By relating these findings to structural considerations of a solution approach represented as an arithmetic tree, statements can be made about its cognitive complexity. On the one hand the arithmetic tree-like structure (Figure 1) can be interpreted as causal chain since the start (given information in the task text) and end (solution of the task) is linked by chain links (intermediate steps in the solution process). On the other hand direct predecessors can be identified as relevant intermediate steps. Thus, it can be concluded that the previous intermediate step must be kept active in the working memory to master the following. The assumption that the mental processing capacity is limited (Sweller, 1988) leads to the statement that several information which has to be kept active at the same time, complicate the solution process. Thus, it can be deduced that the load of the working memory is dependent on the number of intermediate steps necessary to master the current intermediate step. That means that the load of the working memory increases with increasing number of information needed at the respective point in the solution process.

Based on these considerations a study has been performed where theoretical difficulty of a solution approach is characterized as its cognitive complexity (Reit, 2016). Starting point is the so called thought structure of a solution approach which can be interpreted as kind of arithmetic tree (Figure 1). To formulate a thought structure all student solutions of a modelling task have been clustered into several solution approaches according to the mathematical model or solution process used. An
aim of the study was to investigate whether the number of sequential and parallel thought operations has an effect on the cognitive complexity and thus, the theoretical difficulty of a solution approach.

**Study design**

Approximately 1800 grade 9 students (15 years of age) from German grammar schools took part and completed a booklet consisting of three out of five modelling tasks (see modelling task “potato” in Figure 2) under seatwork conditions. The total processing time for a booklet lay within one teaching unit.

![Diagram of potato with length measurement](image)

Industrial manufactured French fries are supposed to be equal in size and the single sticks are cut out lengthwise. Therefore not the whole potato can be used. The potato tubers look similar to the picture above, are regularly formed and approximately 10 cm in length.

How many of these potato sticks can be obtained from one potato? Reason mathematically.

**Figure 2: Modelling task “potato” (Reit, 2016)**

**Method**

In the following the method in the study of Reit (2016) will be briefly explained. For a detailed description of the methodical implementation it is referred to Reit (2016).

All student solutions of a modelling task were analyzed and different solution approaches could be identified (two to four solution approaches per modelling task). These solution approaches within one modelling task differed in their underlying mathematical model used or, if similar to this, in their solution process. Every student solution was finally assigned to a solution approach (Figure 3).

A structural analysis of these solution approaches then lead to individual thought structures indicating the chronology of thought operations. Based on thought structures of solution approaches different difficulty models have been developed to translate the respective structure into a scalar difficulty value. In a first step a thought structure was mapped onto a so called *thought structure vector* (Figure 3). These thought structure vectors represent the tabular-compact form of a thought structure. Each vector component indicates the number of parallel thought operations on the respective level of the thought structure.

Due to the fact that it was not clear yet if parallel thought operations lead to a higher difficulty than sequential thought operations, different operationalization of a thought structure vector into a scalar value were imaginable. Therefore different difficulty models have been set up (four accounting for parallelism of thought operations by weighting them and one non-weighting model (addition model)) which lead to solution approach specific difficulties.
Figure 3: Identification of solution approaches, setting up thought structures (together with its thought structure vector) and applying difficulty models which lead to theoretical difficulties

Results

Whether and to what extent parallel and sequential thought operations have an influence on the difficulty was evaluated by comparison with the corresponding empirical difficulty, as a measure of the average score of a solution approach. To determine the empirical difficulty all student solutions have been assessed by two independent raters on the basis of a predefined assessment scheme set up by experts. The question was whether the theoretical difficulty reflected the associated empirical difficulty of a solution approach. In this case structural characteristics of a solution approach can be taken as a basis for assessment of modelling tasks. Of special interest are the results of the non-weighting difficulty model as analogy to mathematics teachers’ intuitive assessment practice. The non-weighting difficulty model (addition model) adds up all thought operations to arrive at a theoretical difficulty as it is done, more or less intuitively, by mathematics teachers when identifying scoreable intermediate steps.

The results (Figure 4) indicate that addition and factorial model (pseudo-$R^2=0.83$) map the coherence of theoretical and empirical difficulty best. The factorial model weights parallel thought operations. In regard of the focus of the paper the results in Figure 4 clearly show that the addition model lead to significantly better results than the most weighting models.
Discussion of results

Particular reference is made to established but so far not researched assessment practices in mathematics teaching. The focus is on whether intuitive assessment practices of mathematics teachers can be empirically confirmed and transferred to modelling tasks. Intuitive assessment practice means the common procedure of mathematics teachers of scoring intermediate steps in student solutions without accounting for structural-cognitive particularities. These intermediate steps usually then serve as a basis for an assessment scheme. The portrayed procedure is commonly used when assessing performance tasks in mathematics (summative assessment). Modelling tasks with its multiple solution approaches are considered to be hard to assess. This problem not least leads to the fact that modelling tasks are sparsely used in mathematics class. Results of a study investigating the difficulty of modelling tasks support the intuitive assessment practice of mathematics teachers and thus, legitimate transferring this assessment practice to modelling tasks.

In detail the results show that the addition model which treats sequential and parallel thought operations equally lead to reasonable results. This indicates that difficulty of a solution approach can be described well by the number of thought operations needed to arrive at a solution. By applying the addition model it is assumed that parallelism of thought operations has no influence on
the complexity of a solution approach. This assumption is also made, more or less intuitively, in common assessment practice in mathematics teaching. Intermediate steps are identified and scored. Thus, on the one hand the results support the everyday procedure in mathematics teaching where the difficulty of a mathematics task is often interpreted as the number of intermediate steps to complete a solution. On the other hand the results might justify a similar assessment procedure when assessing modelling tasks. In view of the widespread problems concerning assessing modelling tasks in everyday school life as part of the grading, the results clearly highlight a possible and furthermore practicable way.

In summary it can be concluded that the so far intuitive assessment practice in school of identifying intermediate steps in a solution, can be supported by empirical findings. It can be a worthwhile procedure to identify thought structures as a basis for assessment especially when assessing modelling tasks. By assuming that assessment is connected with difficulty of the respective solution approach, parallelism of solution approaches has an influence (see the results of the factorial solution) but might be neglected in favour of a straightforward applicability in everyday school practice. Thus the results of Reit (2016) can serve as a basis for the development of a manageable assessment scheme for modelling tasks and might promote their implementation in mathematics teaching.

References


