Studies in Educational Evaluation 58 (2018) 37–50

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

A quantitative methodology for analyzing the impact of the formulation of a mathematical item on students learning assessment

Giorgio Bolondi[a], Laura Branchetti[b,*], Chiara Giberti[c]

[a] Department of Education, Free University of Bolzano, Viale Ratisbona, 16, 39042, Bressanone-Brixen (BZ), Italy
[b] Department of Mathematics, Physics and Computer science, University of Parma, Parco Area delle scienze, 7/a, 43124, Parma (PR), Italy
[c] I.C. "Fabriani", Viale Marconi, 6, 41057, Spilamberto, Modena (MO), Italy

ARTICLE INFO

Keywords:
Student evaluation
Evaluation methods
Word problems text formulation
Item response theory

ABSTRACT

In this paper, we present a methodological approach to the investigation of the impacts of text formulation on students' answers in mathematical problem solving-based assessment. After a review of the related literature in Mathematics education and a review of the methodologies used until now to investigate this research issue, we describe in depth our quantitative approach, providing motivations and examples of its statistical relevance and its potentiality in making interesting phenomena emerge, to be interpreted with further qualitative methods. We observed statistically significant evidences of different impacts of the variations on different categories of students (males/females; students with high and low performances in the whole test). The methodology and our preliminary results can inform researchers in mathematics education, teachers and experts in the agencies that are responsible for large-scale students learning assessment in several contexts (national and international).

1. Introduction

When facing a mathematical task, students are influenced by the formulation of the task itself. In particular, this may influence in a significant way their performance when dealing with an assessment task, for instance in the case of a word problem. This is a classical topic in educational research; for instance, a recent literature review for the case of arithmetical word problems is Daroczy, Wolska, Meurers, and Nuerk (2015).

A better understanding of the relationships between formulation of a problem/task, reading and problem-solving strategies and students' performances may have three kinds of impact:

- a *theoretical* one, in the direction of problematizing the relation between students' knowledge and the assessment based on students' answers to written tests as "final products": it can contribute to better define the summative aspects of assessment;
- a *practical* one: it may help task-assessment designers (teachers, large-scale assessment authors, researchers….) both in well defining the *question intent* and in monitoring different levels of difficulty;
- a *didactical* one: it may help in interpreting students' behaviors when answering to an assessment question, hence, it can also give a contribution to formative assessment.

In this paper, we propose our methodological contribution to this general problem by designing and validating a quantitative methodology for measuring the impact of a variation in the formulation of an item on students' performances. In particular we present: our background; the steps and the kind of data necessary to carry out a research based on this methodology; a validation plan of the methodology based on the confirmation in two cases of results that we consider "solid findings" in Mathematics education (Education Committee of the EMS, 2011) concerning the impact of formulation in mathematical problem-solving; two examples of quantitative results that put new light on those findings and may encourage researchers in Mathematics education to carry out further qualitative researches on new categories of phenomena that have not been investigated yet since the methodologies used to address the research questions did not take care of such aspects.

The structure of the paper is as follows.

In Section 2, we outline the background of the problem. We present in §2.1 different approaches to the categorization of the variables in the formulation of a mathematical task and some research on the impact of different formulations on the performance of students. This review will allow us in §4 to frame the cases that we will use for the validation of our tool. In §2.2 we review the different methodologies used in the research on the impact of variation of formulation of a task, presenting the main methodological difficulty and showing the lack of a quantitative method for measuring this impact, hence the rationale for this

* Corresponding author.
E-mail addresses: giorgio.bolondi@unibz.it (G. Bolondi), laura.branchetti@unipr.it (L. Branchetti), chiara.giberti@unitn.it (C. Giberti).

https://doi.org/10.1016/j.stueduc.2018.05.002
Received 9 January 2018; Received in revised form 22 April 2018; Accepted 11 May 2018
Available online 26 May 2018

paper. In §2.3 we present the statistical background of our methodology, which is indeed the set of techniques largely used in large scale assessments. §2.4 contains our research questions.

Section 3 is devoted to the description of our method. §3.1 contains the design of the tool, §3.2 the outputs, and §3.3 the coherence and compatibility conditions that must be satisfied in order to consider acceptable the data obtained.

Section 4 contains the validation plan and its results, which is based on a starting test (coming from a large-scale assessment) for which solid data are already available, and variations of formulation for which extensive didactic researches have been already performed. We stress since now the fact that our purpose is not to interpret data in order to provide new results at this stage. Our validation strategy relies on showing how our tool provides data both confirming previous results (obtained with different methodologies), and pointing out new phenomena. In §4.1 we describe our starting test, the variations, and the population. In §4.2 we verify that our test satisfies the coherence and compatibility conditions 3.3, and that our experiment provides general data coherent with the solid data of the original large-scale assessment. In §4.3 we present the output data for two cases and we discuss them under the light of existing didactic research, showing what our method may provide for a quantitative framing of the phenomena. In particular, we show how it highlight new phenomena.

Section 5 contains our conclusions, with our remarks about the future issues and the limit of our approach.

2. Background of the research and statement of the problem

2.1. Variables in the formulation of a problem

We present here the context where our analysis takes place. This review of previous researches will furnish the didactic variables for our study.

During the last decades, many authors studied and classified possible formulations of the test of a mathematical task. Others, inquired about the impact of differences in the test impact on students' behavior. We report in this section some relevant results that we used as solid findings to check with our methodology and, in particular, we describe two results that we analyzed in depths, as we report in the Section concerning the data analysis.

As pointed out by Bagni and D'Amore (2005), the crucial point concerning variations in the formulation is not the fact that a formulation is necessarily better or worse than another one, but the fact that changing the formulation actually changes the problem.

The factors influencing students' approach to the answering of a written test may be a lot and it is complex to list them being exhaustive. However, some attempts have been done to list categories of such factors in the field of mathematics education and we started from them to have a picture of what could be interesting to investigate while facing the problem of measuring the impact of variations in the formulation on students' performances. Analyzing the factors affecting the problem-solving activities, Nesher (1982), while categorizing the variations, listed three components that may vary in a *word problem*: logical (operations, lack or abundance of data, …), syntactic (position of the question in the text, number of words, …) and semantic (contextual relations, implicit suggestions, …).

Considering the more general problem (not necessarily bounded to a mathematical word problem or to arithmetical contents) of the comprehension of a text and the information retrieval, Duval (1991) studied what he called "variables rédactionnelles" (French original name), stating that they influence the student´s cognitive and operative processes. As D'Amore (2000) highlighted, these modifications in the text, even small, may cause changes in the students' approaches to problem solving. Laborde redefined them in 1995 in order to include also other non-verbal variations, such as introducing form of representations (Laborde, 1995). She listed factors concerning editing, punctuation, syntactical complexity, word density, order of the information, explicit declaration of intermediate objects needed for the solution. However, how individuals come up with mathematical solution strategies can also be influenced by numerical factors like number magnitude (Thevenot & Oakhill, 2005). This result is confirmed and analyzed in depths by De Corte, Verschaffel, and Van Coillie (1988), with a focus on the number type (integer, decimal bigger than 1, decimal smaller than 1) in arithmetical word problems concerning multiplication, stressing the difference in students' answers when the number type change in the multiplier, while they stressed that there were no significant changes when the multiplicand changed. We use this as first solid finding to analyse in depths to validate our methodology since it is very detailed and strong from the methodological point of view, ans it has been also mentioned as a solid finding by Daroczy et al. (2015) when they proposed a review of the factors affecting the difficulty of word problems and described the "three components of WP difficulty: (i) the linguistic complexity of the problem text itself, (ii) the numerical complexity of the arithmetic problem, and (iii) the relation between the linguistic and the numerical complexity of a problem". Yet, Daroczy et al. (2015) stated that variations in problem solving strategies could depend on linguistic factors like wording, semantic categories and propositions. The influence of linguistic factors on Mathematics teaching and learning is a classical topic in Mathematics education – see for instance the review by Schleppegrell (2007), that has been investigated a lot in the case of problem solving. A detailed analysis of *word problems* that is relevant from this point of view has been carried out by Frank, Koppen, Noordman, Vonk, and Perfetti (2007). According to the authors (p. 2): "A broad model of text comprehension should not only simulate how information is extracted from the text itself, but also how this information is interpreted in light of the reader's knowledge." This distinction is related to the distinction among three levels of discourse representation: the first level is the *surface representation*, "consisting of the text's literal wording"; the second level "called the *textbase*, where the meaning of the text is represented as a network of concepts and propositions from the text […] connection relations between propositions in a coherent text base are typically expressed by connectives"; the third level of representation, named *situation model*, "textbase elements are combined with elements from the reader's general knowledge".

Branchetti and Viale (2015) contributed to the general statement that linguistic variations can affect students' performances. In particular, they investigated the impact of variations in the syntactic structure of the sentences (variating thus the first and the second level) and highlighted effects of these variations also on students with good performances in mathematics. We referred to the general solid findings concerning the impact of variations of the linguistic factors reported by Daroczy et al. (2015), trying to investigate from a quantitative point of view the statement by Branchetti and Viale (2015) about the students with good performances, comparing two cases that our methodology showed to be very different from the point of view of the students answers distribution

2.2. Research methodologies

The methodologies used in order to investigate the impact of these variations are almost quantitative and often consist in the administration of different tests, containing two or more formulations of the "same" task. In some studies, the same question is revised and reformulated in many versions and all the different forms of the task are administered to the same group of students (e.g. Lepik, 1990; Cummins, Kintsch, Reusser, & Weimer, 1988; De Corte, Verschaffel, & De Win, 1985; Thevenot, Devidal, Barrouillet, & Fayol, 2007). In this case, the ability of the students responding to the different versions of each task is the same but the main problem of this approach consists in the unavoidable influence of the work performed by the student on the first task administered, on his resolution of the second one. In almost all of these researches, the way to partially overcome this obstacle consists in