Journal of Mathematical Behavior 40 (2015) 88–105

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

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

The assessment of mathematical literacy of linguistic minority students: Results of a multi-method investigation

Wolff-Michael Roth[a,*], Kadriye Ercikan[b], Marielle Simon[c], Romeo Fola[c]

[a] University of Victoria, Canada
[b] University of British Columbia, Canada
[c] University of Ottawa, Canada

ARTICLE INFO

Article history:
Available online 10 February 2015

Keywords:
Mathematics achievement
Linguistic minority students
Differential item functioning
Expert analyses
Mathematical cognition
Language differences
Think-aloud protocols

ABSTRACT

Assessing mathematical literacy of students who have limited proficiency in the language of the test is a critical challenge in mathematics education. Previous research indicates that knowledge and competencies of such students are underestimated. This presents a major validity and fairness problem for assessment. Most efforts addressing fairness and validity issues in assessment of linguistic minority students focus on the test language only. To overcome limitations of single approaches, we examine in this study the interaction between the test language and the student language background by means of multiple methods. Thus, we investigate possible linguistic bias of items flagged as functioning differentially (the result of DIF analyses) by means of (a) two levels of expert analyses and (b) student think-aloud protocols to investigate language effects in published mathematics items from the 2000 to 2003 Programme for International Student Assessment (PISA) administration for students attending French schools in Canada and speaking either French or other languages at home. DIF analyses were conducted to identify items on which students from different home language backgrounds attending French schools achieve differently. The expert panels tended to identify surface characteristics of language that may be responsible for group differences but not for the differential effects detected by differential item functioning (DIF). Student think-aloud protocols in part confirm and in part contradict DIF results, providing insights for the source of the differences. Suggestions are provided for further study.

1. Introduction

Every three years, the results of the most recent international survey Programme for International Student Assessment (PISA) are both eagerly awaited and dreaded. While this study was conducted, one of the leading national newspapers in Canada calls attention using the title "Canada's fall in math-education ranking sets off alarm bells" (Alphonso, 2013). The article not only reports that Canada has dropped out of the top 10 "placing 13th overall, down three spots from 2009 and six spots from 2006." The article further suggests that mathematics scores were dropping not only in Canada but also worldwide and that "the math curriculum, ushered in over the past decade, is to blame for lower scores because it places more emphasis on real-world concepts rather than abstract thinking and practice." This Canadian reaction is reflected

* Corresponding author. Tel.: +1 250 721 7764.
E-mail address: wolffmichael.roth@gmail.com (W.-M. Roth).

http://dx.doi.org/10.1016/j.jmathb.2015.01.004
0732-3123/

in many other industrialized (Organisation of Economic Co-operation and Development, OECD) countries as well, where mathematics education is either criticized or, if the country has improved in the rankings (e.g., Germany), the distance to the achievement scores of the leading countries is noted as a major concern (e.g., Friedman & Trenkamp, 2013). In France, the most recent results constituted a shock, and "poor showings" have led to reflections concerning the inequities of an educational system where the proportion of elite mathematics students remained the same over the past few years, but where the proportion of students in difficulties (e.g., from disadvantaged families) has risen by a factor of 1.35 (Battaglia & Collas, 2013). Even though there are cautions about the over-interpretation of PISA and other international assessments in education generally (e.g., Ercikan, Roth, & Asil, 2015) and in mathematics education specifically (Brown, 2012; Tsatsaroni & Evans, 2013), their influence on national mathematics education policy is considerable (e.g., Boesen et al., 2014) and therefore requires continued scholarly engagement. This is so because the test results "are employed to make significant resource allocations, curriculum planning, and strategic decisions" (Roth, Oliveri, Sandilands, Lyons-Thomas, & Ercikan, 2013, p. 547).[1]

Although the results of the latest PISA assessments tend to make headlines, questions concerning the validity and limitations of the inferences based on it are much more rarely discussed (e.g., Leung, 2014), including, for example, whether multi-lingual tests comparably assess competencies in different language groups—e.g., French and English version of a pan-Canadian test (Ercikan, Roth et al., 2015)—or in the same language but for students living in minority situations (Ercikan et al., 2015) or English language learners contexts (e.g., Ercikan, Roth, Simon, Lyons-Thomas, & Sandilands, 2014). Thus, one recent study has shown that 19, 21, and 25% of the reading, mathematics, and science PISA items had different psychometric properties for students from English and non-English language backgrounds attending English-language schools in Australia, Canada, the UK, and the US (Ercikan et al., 2015). Another study compared PISA mathematics achievement of Quebec students—who speak French at home, live in a French majority setting, and attend schools with French as instructional language—and Ontario students, who attend French-instruction schools in a minority setting and speak (OFF) or do not speak French at home (OFNF) (e.g., Ercikan et al., 2014). That study showed that of 163 mathematics items across the three assessments conducted between 2000 and 2006, 22 items (13.5%) exhibited bias against one or another group.

The purpose of this research is to investigate possible linguistic bias in PISA items differentially solved by students from different language groups. Most efforts addressing fairness and validity issues in assessments of linguistic minority students focus on assessment language only (Au, 2013; Abedi, 2004; Abedi, Hofstetter, & Lord, 2004; Abedi & Lord, 2001; Butler, Bailey, Stevens, Huang, & Lord, 2004; Nguyen & Cortes, 2013; Vale et al., 2013). In this research we use a mixed-method approach to examine the interaction between the test language and the student language background on large-scale, international examinations with a particular focus on the released PISA 2000 and 2003 mathematics items using four levels of analyses: (a) item-response theory (IRT) based analyses of differential item functioning (DIF) for the selection of items included in analyses (b) through (d), (b) analyses by experts in educational measurement and mathematical cognition, (c) language and curriculum experts, and (d) think-aloud protocols (TAPs) with 33 students from the target populations. DIF analyses help us identify items where students from different language backgrounds have differential response patterns. But this statistical approach does not tell us whether and how language may be at play in the differential response patterns. The expert reviews are the first step to understand if item language may be at play. TAPs help us examine if and how student language may be the source of DIF.

2. Background

2.1. Possible language bias in large-scale assessments

The validity of comparisons of scores from large-scale assessments, such as PISA critically depends on whether the tests actually assess the same or similar knowledge and competencies across the groups and sub-groups that are compared (Hambleton, Merenda, & Spielberger, 2005). Most research concerning bias in large-scale, international assessments focuses on bias arising from incomparable constructs (tests measure different forms of knowledge and competencies) or measurement (the construct is the same but measurement is biased). In this study we investigate possible sources of bias for different language groups: students who attend schools with French as the language of instruction situated in English as the dominant language (Ontario) context and who speak or do not speak French at home. The latest PISA report suggests that in mathematics, "Canadian students continue to perform well in mathematics in a global context" (Brochu, Deussing, Houme, & Chui, 2012, p. 18) scoring 24 points above the OECD mean, being outperformed by only nine countries. In Quebec, where many students take the test in French, the mean score was 18 points above the Canadian mean, whereas in Ontario, where most students take the test in English, the students achieved 4 points below the national mean. In fact, Quebec was the only province where students scored above the national mean; at the same time, in this province, there was the largest difference between low and high achievers.[2] The difference between high and low achievers was lower than the national average in the

[1] The relevant level differs according to jurisdiction. In Canada and Germany, cultural authority and educational decision-making lies with the provinces (Länder). In countries with central governments, such as France, educational decisions are made nation wide.

[2] This difference is due to the larger proportion of high achievers; and it is observed in high-achieving countries generally (Brochu et al., 2012). The difference does not mean, however, that there is an issue with equity.