Learning and Instruction 28 (2013) 24–34

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

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

# Second language learners' performance in mathematics: Disentangling the effects of academic language features

Nicole Haag [a,*], Birgit Heppt [b,1], Petra Stanat [a], Poldi Kuhl [c,2], Hans Anand Pant [a]

[a] Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany
[b] Department of Educational Science and Psychology, Educational Research, Freie Universität Berlin, Habelschwerdter Allee 45, D-14195 Berlin, Germany
[c] Berlin-Brandenburg Institute for School Quality Improvement (ISQ), Otto-von-Simson-Str. 15, D-14195 Berlin, Germany

ARTICLE INFO

*Article history:*
Received 19 May 2012
Received in revised form
16 April 2013
Accepted 17 April 2013

*Keywords:*
Educational measurement
Second language learners
Differential item functioning
Academic language
Test validity and fairness

ABSTRACT

Several studies have shown that the linguistic complexity of items in achievement tests may cause performance disadvantages for second language learners. However, the relative contributions of specific features of linguistic complexity to this disadvantage are largely unclear. Based on the theoretical concept of academic language, we used data from a state-wide test in mathematics for third graders in Berlin, Germany, to determine the interrelationships among several academic language features of test items and their relative effects on differential item functioning (DIF) against second language learners. Academic language features were significantly correlated with each other and with DIF. While we found text length, general academic vocabulary, and number of noun phrases to be unique predictors of DIF, substantial proportions of the variance in DIF were explained by confounded combinations of several academic language features. Specialised mathematical vocabulary was neither related to DIF nor to the other academic language features.



## 1. Introduction

In Germany, the number of students with an immigration background has constantly increased over time. In 2009, almost 26 per cent of 15-year old students in German schools had at least one foreign-born parent (Stanat, Rauch, & Segeritz, 2010). Providing valid and fair assessments for this growing and heterogeneous group of students is among the major challenges in national and international large-scale assessments. Group differences, such as differences in math achievement between students who primarily speak German at home and students who primarily speak a language other than German at home, can only be meaningfully interpreted if the underlying tests are equally valid for both groups. However, concerns have been raised about the validity of standardised achievement tests for second language learners (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). If students are tested in a language that is not their native language, the scores might reflect not only their competencies in the measured content area (e.g., mathematics), but also their mastery of the language. Math problems entailing linguistic demands are expected to impede second language learners from fully understanding the items and, hence, from demonstrating their mathematical ability. Native speakers, in contrast, are more likely to possess the language skills necessary for understanding linguistically challenging math word problems (cf. Wolf et al., 2008). Thus, items with high linguistic demands (e.g., long and complex instructions) in mathematics assessments may measure construct-irrelevant language competencies (Martiniello, 2009; Wolf et al., 2008). These items could be disproportionally more difficult for second language learners and, thus, less appropriate for capturing their mathematical competencies than for native speakers of the respective language.

On the item level, this difference in difficulty can be detected with differential item functioning (DIF). The presence of DIF indicates differing item difficulties for two groups of students with the same ability level (Holland & Wainer, 1993). Large DIF values indicate construct-irrelevant variance in the test scores and suggest that the item, and hence the test, may be differentially valid for the subgroups. In order to ensure equal validity for native speakers and second language learners, it is crucial to determine the particular language features that may potentially contribute to construct-

* Corresponding author. Tel.: +49 30 2093 46578; fax: +49 30 2093 46598.
*E-mail addresses:* nicole.haag@iqb.hu-berlin.de (N. Haag), birgit.heppt@hu-berlin.de (B. Heppt), petra.stanat@iqb.hu-berlin.de (P. Stanat), poldi.kuhl@iqb.hu-berlin.de (P. Kuhl), hansanand.pant@iqb.hu-berlin.de (H.A. Pant).
[1] Now at the Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
[2] Now at the Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany.

0959-4752/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.learninstruc.2013.04.001

irrelevant difficulty for second language learners and, thus, cause DIF (e.g., Abedi, Lord, & Plummer, 1997; Martiniello, 2009).

### 1.1. Academic language

Various researchers have argued that academic language differs fundamentally from everyday language (Bailey & Butler, 2003; Butler, Bailey, Stevens, Huang, & Lord, 2004; Cummins, 1979, 2000; Schleppegrell, 2004). Accordingly, Cummins (1979) differentiates between *Basic Interpersonal Communication Skills (BICS),* which are acquired in everyday interactions, and *Cognitive Academic Language Proficiency (CALP),* which is acquired in the context of schooling and is essential for understanding cognitively demanding contents in context-reduced learning settings. While second language learners may understand and participate in casual conversations in everyday life quite well, they are expected to have more problems with academic language, which makes it difficult for them to fully understand classroom discourse as well as academic readings and tasks. This certainly does not mean that that social language is inherently or generally less sophisticated than academic language, yet the two language registers are assumed to differ in certain aspects, particularly "in the relative frequency of complex grammatical structures, specialized vocabulary, and uncommon language functions" (Bailey, 2007, p. 9; see also Schleppegrell, 2001, 2004).

An integrative conceptualization of academic language that can be used to describe the language features of test items was developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST; see Bailey & Butler, 2003; Butler, Lord, Stevens, Borrego, & Bailey, 2004). Based on the theoretical assumptions described above, the authors conceive of academic language as the language that is spoken in the classroom or other academic contexts in order to impart and acquire knowledge (Chamot & O'Malley, 1994). To capture characteristics of academic language in texts and instructions of test items, they can be rated in terms of various descriptive, lexical, and grammatical criteria (Bailey, 2000/2005; Bailey, Butler, Stevens, & Lord, 2007; Butler, Bailey, et al., 2004). *Descriptive features* include the overall number of words and average sentence length. These features are related to reading difficulty (Zakaluk & Samuels, 1988), with longer sentences posing greater challenges for readers (Butler, Bailey, et al., 2004; Groeben & Christmann, 1996). *Lexical features* encompass general academic vocabulary, which is used across school subjects and disciplines (e.g., "report", "synthesize"), as well as specialised academic vocabulary, which is associated with a specific discipline (e.g., "denominator", "multiplication"). These words are usually abstract and semantically opaque (e.g., Corson, 1997; Townsend & Collins, 2009). *Grammatical features* that tend to appear in academic contexts more often than in everyday language are mainly verb forms in passive voice, prepositional phrases, noun phrases, and participial modifiers. Dehn (2011) and Gogolin (2004) assume that in German, grammatical features, such as compound sentences and long and complex noun phrases, may play a more important role than lexical features. However, empirical evidence for this assumption is scarce.

Furthermore, it has been hypothesised that the impact of specialised academic vocabulary on second language learners' performance in academic content domains should be less pronounced than the impact of the other academic language features. Some researchers suggest that academic language in general and specialised academic vocabulary in particular are, to some extent, acquired in different contexts (e.g., Gogolin, 2003). Specialised mathematical academic vocabulary is often explicitly introduced and explained in mathematics instruction (e.g., Ernst-Slavit & Mason, 2011; KMK, 2004), whereas general academic words are rarely explained in classroom discourse (cf. Komor & Reich, 2008). This may be particularly true for the German language in which general academic vocabulary is sometimes derived from vocabulary used in everyday communication (see Appendix for examples from this study). In the school setting, these words are associated with somewhat different meanings or used in different contexts (Grießhaber, 2010). Accordingly, German linguists (e.g., Ehlich, 1999) have developed the concept of everyday academic language ("Alltägliche Wissenschaftssprache") which refers to everyday vocabulary that has "undergone a functional change" (Skrandies, 2011, p. 100) and is now used in new collocations in school-related contexts. Given the superficial resemblance between everyday vocabulary and this specific part of general academic vocabulary, teachers are likely to assume that students understand the terms used in the classroom and do not deliberately and systematically explain them. Therefore, students' mastery of academic language strongly depends on their opportunities to acquire academic vocabulary outside school, which specifically disadvantages second language learners, as they tend to have a smaller knowledge base on which they can draw to infer the meaning of general academic words than native speakers (e.g., Dubowy, Ebert, von Maurice, & Weinert, 2008). They are thus likely to lag behind their native peers in their command of academic language (cf. Gogolin, 2003). The differences between second language learners and native speakers should therefore be larger for general academic vocabulary than for specialised academic vocabulary.

Second language learners with limited knowledge of academic language may also be hindered in creating mental models of mathematical problems (cf. Heinze, Herwartz-Emden, & Reiss, 2007). Presumably, these students experience a higher amount of cognitive load when confronted with linguistically demanding mathematics test items than their monolingual peers (Campbell, Adams, & Davis, 2007; see also Paas, Renkl, & Sweller, 2003). As working memory capacity is limited, their performance might suffer from the additional load associated with the presence of academic language features in mathematics items.

In the past, two types of studies have been performed to determine the relationship between language features of test items and performance of second language learners in content domain assessments, such as mathematics and science. Most of these investigations were conducted in English-speaking countries. The first type of studies compared the performance of native speakers and second language learners on test items with different levels of language demands. The second type of studies analysed differences in item difficulty between second language learners and native speakers (DIF). In studies of the second type, test items that were found to show DIF against second language learners were analysed in terms of their respective language features. Most studies identified these language features in post-hoc-analyses by simply comparing the linguistic features of items with and without DIF. One study predicted DIF with a composite score of linguistic complexity, yet it did not explore the unique effects of different language features. An overview of the results of both types of studies is provided in the following two sections.

### 1.2. Relationship between academic language features and mathematics performance for second language learners and native speakers

In the first type of studies described above, most analyses compared the performance of students who learn English as a second language – usually referred to as English language learners (ELLs) – with the performance of proficient speakers of English. Some of these studies suggest that ELLs score lower on mathematics items containing longer item stems than on language-free