University of Groningen

The relation between student's math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education

Korpershoek, Hanke; Kuyper, Hans; van der Werf, Margaretha

Published in:
International Journal of Science and Mathematics Education

DOI:
[10.1007/s10763-014-9534-0](https://doi.org/10.1007/s10763-014-9534-0)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Korpershoek, H., Kuyper, H., & van der Werf, M. (2015). The relation between student's math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education. *International Journal of Science and Mathematics Education*, 13(5), 1013-1037.
<https://doi.org/10.1007/s10763-014-9534-0>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

THE RELATION BETWEEN STUDENTS' MATH AND READING ABILITY AND THEIR MATHEMATICS, PHYSICS, AND CHEMISTRY EXAMINATION GRADES IN SECONDARY EDUCATION

Received: 19 September 2013; Accepted: 12 March 2014

ABSTRACT. Word problems are math- or science-related problems presented in the context of a story or real-life scenario. Literature suggests that, to solve these problems, advanced reading skills are required, in addition to content-related skills in, for example, mathematics. In the present study, we investigated the relation between students' reading ability and their achievements in advanced mathematics, physics, and chemistry when controlling for their mathematical ability. The study included 1,446 Dutch secondary school students who had taken their final examinations in these subjects. Using multivariate multilevel models, we found that both math and reading ability (the latter only in the pre-university track) were positively related to the examination grades on mathematics, physics, and chemistry. This relation was equally strong for boys and girls. This relation was neither moderated by students' sex, nor by students' ethnicity (Dutch versus non-Dutch origin).

KEYWORDS: math ability, multivariate multilevel models, reading ability, secondary education, word problems

Word problems (or “story-problems”) are math- or science-related problems presented in the context of a story or real-life scenario (Adams, 2003, p. 790). In secondary school books, they are commonly used in addition to more traditional problems that have to be solved by correctly applying mathematical, physics, or chemistry formulae. They generally contain textual instructions or a situational context for finding a solution (Chapman, 2006; Lantz-Andersson, Linderoth & Säljö, 2009). Word problems require students to move between two symbolic codes: written language and mathematical symbols (Lantz-Andersson et al., 2009). Word problems largely differ in their reading difficulty, for instance, in terms of number of words used, word familiarity, or word difficulty. They require the identification of relevant details and relationships between concepts (i.e. reading comprehension), which are often hidden within the language. To solve a word problem, the students must derive the relevant information from the text to set up a calculation problem for retrieving the missing information (Fuchs, Fuchs, Fletcher, Hamlett & Lambert, 2008).

Evidently, advanced reading skills are needed to deal successfully with a number of learning tasks, such as solving contextualized problems, interpreting written assignments, and following the oral or written instructions given by the teacher. Contextualized problems are frequently used in mathematics, physics, and chemistry examination tests in secondary education. It is therefore plausible that some students do not perform well on these tests because their reading skills inhibit their ability to solve these contextualized problems, even though they have a good understanding of course content itself (see also Reusser & Stebler, 1997).

In this paper, we examined the significance of students' reading ability for their academic achievement in mathematics, physics, and chemistry in secondary education. As it is well-known that students' achievement in these subjects is partly predicted by their math ability, we included a measure of students' math ability as well. Using both ability measures, we sought to understand to what extent students' reading ability is related to their academic achievement in mathematics, physics, and chemistry when taking students' math ability into account. Moreover, we examined whether this relationship was moderated by gender and ethnic background (Dutch versus non-Dutch origin). To our knowledge, the impact of students' reading ability on mathematics, physics, and chemistry understanding has not been studied extensively.

THEORETICAL FRAMEWORK

In this theoretical framework, we first explain the use of word problems in mathematics- and science-related courses in secondary education in more detail. Second, the relation between reading and math/science is clarified. The focus in most studies has been on the relation between reading and mathematics rather than reading and physics or chemistry; therefore, the general focus of this paragraph is limited to the former relation. Following the reported findings in the literature, we finally present the aims of the present paper.

Word problems in mathematics- and science-related courses in secondary education

There is an ongoing concern that a large part of the test items in math- and science-related examinations assess students' reading comprehension rather than (or next to) their subject understanding (Flick & Lederman, 2002). Reading literacy is a broader conceptualization of reading comprehension; it is generally conceived as the ability to understand

and use written documents containing both verbal and pictorial information, for example texts, pictures, charts, and tables (Baumert, Lüdtke, Trautwein & Brunner, 2009, p. 167) while mathematical literacy is the ability to mathematize real-world situations and appropriately use mathematics in problem contexts (OECD, 1999). The conceptualization of mathematical modeling in mathematics, physics, and chemistry courses goes far beyond the mere computation and application of formulae (Baumert et al., 2009). In a similar vein, Bruner (1986) distinguishes two modes of knowing—paradigmatic and narrative—as two irreducible approaches to cognitive functioning when solving word problems. Paradigmatic knowing requires a focus on universal and context-free mathematical models or structures while narrative knowing demands an emphasis on the social context of the problem. This social context could involve a cover story of the word problem based on which (among other issues) the storyline, situations, and relationships can be understood. The textbooks of mathematics, physics, and chemistry courses in secondary education often use these contextualized problems as a tool for introducing the ‘real world’ into the classroom (see also Ogborn, 1996). Particularly with respect to narrative knowing, sufficient reading skills are required to understand the word problem presented. However, Chapman (2006) reports that secondary school teachers generally emphasize paradigmatic knowing. Accordingly, she suggests that incorporating both modes of instruction would be more consistent with the current perspectives on learning mathematics (see also Kurth, Kidd, Gardner & Smith, 2002 on students’ use of narrative and paradigmatic forms of dialogue in science conversations).

The reading–math/science relation

The fact that most math and science textbooks require reading skills has led several scholars to investigate the relation between students’ reading ability¹ and their achievements in these courses. Some researchers evidently suggest that students’ reading ability is correlated with both their general school achievements and those in mathematics (Ní Ríordáin & O’Donoghue, 2009; Reikerås, 2006). Reikerås (2006), for example, found that low achievement in reading slightly interfered with students’ development of arithmetic skills. Walker, Zhang & Surber (2008) believe that mathematics items designed to measure higher cognitive skills, such as problem solving and mathematical reasoning, are two-dimensional in that they measure both reading ability and mathematics skills (pp. 163–164). In their study (Walker et al., 2008), they found that students’ scores

on these contextualized items were indeed influenced by their level of reading ability. Furthermore, Grimm (2008) reports that early reading comprehension (third grade) relates to a conceptual understanding of mathematics and the application of mathematics knowledge.

We did find several studies on this topic (mostly conducted in secondary education) which describe how students' comprehension of word problems can be improved (Borasj & Siegel, 1998; Brown & Ryoo, 2008; Carter & Dean, 2006; Helwig & Almond, 1999; Nathan, Kintsch & Young, 1992). This focus suggests the assumption that students' reading ability is related to their understanding of math/science. Brown & Ryoo (2008) demonstrated that students, who were taught via a "content-first" approach facilitating the transition from an everyday understanding of general phenomena to the use of scientific language, significantly improved their understanding of science. As regards the text comprehension in mathematics tests, Helwig & Almond (1999) have suggested a video format by which the questions are read by an actor on a video monitor, a method specifically meant for students with an above average mathematical understanding but low reading skills. Additionally, Nathan et al. (1992) proposed using computer-based tutors to help students improve their situational understanding of algebra word problems. Borasj & Siegel (1998) give examples of how transactional reading strategies can support mathematics instruction. They argue that unfamiliarity with the vocabulary necessary to understand word problems can affect students' performance. In the extensive math/science vocabulary, everyday words such as 'product' and 'volume' take on new meanings. Several scholars have suggested that teachers should explicitly teach their students this vocabulary. Whether all these methods can successfully improve students' understanding of word problems in chemistry and physics courses as well is as yet unclear. In recent years, increased attention has been paid to the role of vocabulary and, for example, analogies in teaching chemistry in secondary education (e.g. Bellocchi & Ritchie, 2011). Mathematics teachers generally do guide their students in improving their reading comprehension of mathematics texts (Carter & Dean, 2006). Notwithstanding the importance of this kind of teacher support, however, Chapman (2006) reports that most teachers have little experience in dealing with context in their teaching (i.e. the narrative mode of knowing; see also Bell, Matkins & Gansneder, 2011 on the impact of contextual instruction on teachers' own understanding of science).

All in all, we found multiple indications in the literature that students' reading ability and their academic achievement in mathematics (and

presumably also their achievement in physics and chemistry) are somehow related.

The present study

Based on the literature review above, we conclude the following. First, whereas the general relationship between prior school achievement and future school results has been addressed by many scholars in the field, research on the presumed connection between reading ability and math ability and academic achievement in mathematics, physics, and chemistry courses in secondary education has until now remained scarce. Most studies on the relationship between students' reading ability and their achievements in mathematics have been conducted in primary education (e.g. Kintsch & Greeno, 1985; Lee, Deaktor, Hart, Cuevas & Enders, 2005; Powell, Fuchs, Fuchs, Cirino & Fletcher, 2009; Stern, 1993; Vilenius-Tuohimaa, Aunola & Nurmi, 2008; Walker et al., 2008).

Second, to our knowledge there are not many studies that relate reading ability to academic achievement in physics and chemistry. The studies we found generally deal with mathematics achievement. Similar to the mathematics textbooks, the course materials used in physics and chemistry also include contextualized (word) problems.² Only few studies, however, have paid attention to the reading difficulty of physics and chemistry textbooks (e.g. Jacobson, 1965; Koch & Eckstein, 1995; Koch & Gunstone, 2001; Yore, 1991) or the text comprehension required for biology texts (Ozuru, Dempsey & McNamara, 2009). An exception is a publication by Fagan (1997), who suggests that there may be an important connection between students' reading ability and their achievements in advanced level physics at the end of secondary school.

Third, the studies on the relation between reading ability and math achievement usually concentrate on students with reading deficiencies (e.g. Kytälä, 2008) or English language learners in English-speaking countries (e.g. Beal, Adams & Cohen, 2010; Lee, 2004; Shaw, Bunch & Geaney, 2010; Stoddard, Pinal, Latzke & Canaday, 2002; Turkan & Liu, 2012), whereas reading ability is evidently important for the academic achievement of all students. It is as yet unclear to what extent reading ability and math ability relate to students' academic achievement in mathematics, physics, and chemistry in general.

The current study addressed these three issues. The present study aims at increasing our understanding of the possible significance of students' math and reading ability for their academic achievements in mathematics, physics, and chemistry in a large student sample in secondary education,

consisting of 1,446 students. Korpershoek, Kuiper, Van der Werf & Bosker (2011) have presented correlations of 0.18 (senior general secondary education) and 0.24 (pre-university education) between the students' math ability and their average examination grades on advanced mathematics, physics, and chemistry examinations, which is why we used math ability as covariate in this study. Tests were administered in the first 3 years of secondary education (seventh to ninth grade) to assess students' reading and math ability, and students' exam scores in mathematics, physics, and chemistry were collected at the end of secondary education (11th or 12th grade). Including measures of both physics and chemistry achievement in addition to math achievement is, to our knowledge, new in the field. Furthermore, we examined whether the stated relationship was moderated by gender and ethnic background (Dutch versus non-Dutch origin). Contextualized education might be unfavorable to some students, for example, to those for whom the language of instruction is not their first language (e.g. Dutch for ethnic minority groups in The Netherlands). Several studies have shown that English Language Learners' reading proficiency is related to their mathematics achievement in secondary education (e.g. Beal et al., 2010; Turkan & Liu, 2012). Furthermore, the Programme for International Student Assessment (PISA) 2006 study (Organisation for Economic Co-operation and Development [OECD], 2007) has revealed that in all OECD countries girls on average outperform boys in reading (see also Mullis, Martin & Foy, 2008). Presumably, boys are generally slightly disadvantaged compared with girls as regards the use of word problems in contextualized education. Whether contextualized education is associated with underperformance among particular student groups is as yet unknown. Based on the literature review, we formulated the following hypotheses:

Hypothesis 1: Students' reading ability is positively related to their examination grades on the subjects advanced mathematics, physics, and chemistry when controlling for their math ability.

Hypothesis 2: The relation proposed is stronger for students from ethnic minority groups than for native Dutch students and stronger for boys than for girls.

We did not have clear expectations for which of the three school subjects investigated (advanced mathematics, physics, or chemistry) the relation was the strongest. Therefore, no hypothesis was included on this topic. The three school subjects were separately analysed to find out which of the courses seemed to be the most 'language-oriented.'

METHOD

Participants

The data used formed part of a large-scale longitudinal study conducted in The Netherlands, the "Cohort Studies in Secondary Education" (VOCL '99). The complete sample is representative for all schools and students in Dutch secondary education (Van Berkel, 1999). For more information on the VOCL'99 study, we refer to Korpershoek, Kuyper & Van der Werf (2006). The present study concentrated on the tracks that prepare students for higher education; these are senior general secondary education (track B; duration 5 years), which prepares for higher professional education, and pre-university education (track A; duration 6 years), which prepares for university. The main difference between the two tracks is the difficulty level of the school subjects, with track A being the most difficult track in Dutch secondary education. The participants in this study started their educational career in secondary education in 1999 (usually at age 11 or 12 years) and took their Final School Examinations (FSE) in the spring of 2004 (track B) or 2005 (track A).

We selected the students who had taken their FSE in advanced mathematics, physics, and chemistry (overall sample, $N=1,752$). Based on the availability of prior reading ability and math ability scores and the students' examination grades, we could include 1,446 students (83 %) in the analyses. This response group consisted of 931 track A students (from 108 different schools) and 515 track B students (from 71 different schools). The socioeconomic status (SES) of the students (based on parental education) was relatively high. Most parents of the students included had pursued higher education (60 %), while 33 % had finished secondary education without further study. Only 7 % of the parents had attended primary education and/or a few years of secondary education (low SES students). For some students, the ability measures and/or final examination grades were unavailable due to the non-participation of schools and/or drop out of individual students during the data collection over time. Some schools, for example, prematurely ended their participation in the cohort study or did not administer all requested tests. A small number of students were not included in all waves of the data collection because they declined to participate or had dropped out of school. The "RESULTS" section also contains the non-response analysis.

Variables and instruments

EXAMINATION GRADES. At the end of the 11th (track B) or 12th (track A) grade, students in The Netherlands take part in the national

examinations (designed by specialists at the Dutch National Institute for Educational Testing [Cito]) and school examinations (developed by the schools themselves). The national examinations are of interest here, because these tests are the same across all schools. The examinations in track A are more complex than those for track B. For each course, students obtain a grade between 1 (lowest grade) and 10 (highest grade) (a 6 is a pass). The grades on the national tests in advanced mathematics, physics, and chemistry among our student sample (retrieved from Statistics Netherlands) varied from 2.5 to 9.9 and had a normal distribution (skewness and kurtosis < 1). For 164 track A students and 94 track B students (both 15 %), the final examination grades were unavailable.

MATH ABILITY. The operationalization of math ability was based on three math-related tests. The combination of these three tests yielded a highly reliable measure of the students' math ability. The first test was an arithmetic test administered in the seventh grade, developed by Cito. The reliability (α) of the test was 0.83. It included 20 multiple choice items that covered the following topics: (1) numbers and calculations, (2) proportions, fractions, and percentages, and (3) measuring, geometry, time, and money. Second, an intelligence test was used (the *Groninger Intelligentietest voor Voortgezet Onderwijs* [the Groningen Intelligence Test for Secondary Education], Van Dijk & Tellegen, 1994). This test, assessed as reliable and valid (Evers, Van Vliet-Mulder & Groot, 2000), was administered in the eighth grade. It consisted of a verbal and a symbolic intelligence part, of which we used the scores on the symbolic component (α for the symbolic intelligence part was 0.93). The symbolic intelligence part covered two topics: numbers (25 items) and filling in symbols (20 items). Third, we used a mathematics test for the ninth grade (also developed by Cito) with a reliability (α) of 0.78. This test contained 33 multiple choice items, covering four topics: (1) arithmetic, measuring, and estimation, (2) algebra/relations/graphs/functions, (3) geometry, and (4) statistics and probabilities. In the overall student group of VOCL'99, the correlations among the three tests were 0.42 (arithmetic–symbolic intelligence), 0.50 (mathematics–symbolic intelligence), and 0.52 (arithmetic–mathematics), respectively. A combined math ability score was calculated for students who had completed at least two of these three tests. This was done as follows. A factor analysis on the three (standardized) test scores resulted in a one factor model with an Eigenvalue of 1,963 (65 % common variance) and communalities of 0.64 (arithmetic), 0.71 (symbolic intelligence), and 0.62 (mathematics). The factor loadings were 0.80 (arithmetic), 0.84 (symbolic intelligence),

and 0.79 (mathematics), respectively. Based on these results and the moderate correlations between the three tests, we concluded that the three tests largely measured one common aspect of math ability but that they also measured some other aspects of math ability (i.e. 35 % unique variance). Consequently, we used weighted coefficients rather than unweighted coefficients to calculate students' math ability, following this procedure. To calculate the math ability scores for all students, we first conducted a regression analysis for the full cases (i.e. that had valid scores on all three tests). The resulting regression coefficients were then used in regression equations to predict students' math ability based on the weighted test scores. For the students who had completed two out of three tests, separate regression analyses were conducted for each combination, each based on the weighted test scores of the test scores that were available. Again, the resulting regression coefficients were used to predict students' math ability. By doing so, the relative weight of each test that the student had taken was taken into account when calculating students' math ability. Finally, the scores were standardized again.

Throughout the overall student sample, the range of this (standardized) math ability score ran from -3.55 up to 4.44 . Following recommendations of Kamata, Turhan & Darandari (2003), we used stratified-alpha (as proposed by Cronbach, Schönemann & McKie, 1965) to estimate the reliability of the constructed variable, which was 0.92. The advantage of using this combined score was that also students who had completed only two of the three tests received a math ability score on the same scale as the students who had completed all three tests, which strongly increased the available sample size for the study. For 22 track A students (2 %) and 19 track B students (3 %), no math ability score was available.

READING ABILITY. The operationalization of reading ability was based on three reading-related tests, following a procedure similar to that of math ability. The first one was a Dutch language skills test administered in the seventh grade, developed by Cito. This test of 20 items covered the topics writing, spelling, text comprehension, and vocabulary. Its reliability (α) was 0.74. Second, we used the students' scores on the verbal part of the intelligence test described above (α for the verbal intelligence part was 0.92) which was administered in the eighth grade. The verbal intelligence part covered three topics: synonyms (30 items), analogies (25 items), and categories (30 items). Third, we employed a Dutch comprehensive reading test (40 items) for the ninth grade (again also developed by Cito) with a reliability of 0.75. The correlations among the tests were 0.38 (language skills–verbal intelligence), 0.38 (comprehensive reading–verbal intelligence), and 0.29 (language skills–compre-

hensive reading). We calculated a combined reading ability score for students who had completed at least two of these three tests. The factor analysis on the three tests resulted in a one factor model with an Eigenvalue of 1.719 (57 % common variance) and communalities of 0.54 (seventh-grade reading skills), 0.64 (verbal intelligence), and 0.53 (ninth-grade reading skills). In general, they measured one common aspect of reading ability, but the separate tests also measured other aspects of reading ability (i.e. 43 % unique variance). Our ability measure is thus based on a slightly broader range of aspects of reading ability (hence, the higher unique variance of the separate tests and the lower correlations among the tests) than was the case with the math ability measure. Throughout the overall student sample, the range of the (standardized) reading ability score ran from -3.51 up to 4.55 . The estimated reliability (stratified-alpha) of the constructed variable was 0.83. For 40 track A students (4 %) and 29 track B students (5 %), no reading ability score was available.

SEX. Data about student students' gender were acquired by a questionnaire administered in the seventh grade. In our sample, 831 students were boys (57 %) and 615 were girls (43 %).

ETHNICITY. The ethnic background of the students was determined by questionnaire data collected among the seventh-grade students' parents. Based on the parents' responses, the students were categorized as either native Dutch students or as students from ethnic minority groups (generally children with parents from non-Western European countries). The sample consisted of 1,286 native Dutch students (89 %) and 156 students from ethnic minority groups (11 %), which is representative of the Dutch population, and four non-responders.

Analyses

We applied two three-level models (one for each track) containing the students responses to our criterion variables nested within students, and the students nested within the schools, to represent the hierarchical structure of the data. The data were analyzed with multivariate multilevel models, using MLwiN 2.23 software. Our criterion variables were the examination grades on mathematics, physics, and chemistry for both separate tracks. Next to unconditional models without predictors, several models containing explanatory variables (math ability, reading ability, sex, and ethnicity) were tested. Finally, we added two-way interaction terms to the models between (1) math and reading ability, (2) the ability measures and sex, and (3) the ability measures and ethnicity.

RESULTS

Non-response analysis

Table 1 gives an overview of the average scores of the response and the non-response groups on the included ability and criterion variables.

In track A, the response group scored on average somewhat higher on reading ability than the non-response group, $t = -3.44$, $df\ 969$, $p < 0.001$ (Cohen's $d = 0.56$). This result also applied to the physics grades, $t = -2.43$, $df\ 1,093$, $p < 0.05$ ($d = 0.23$), indicating that in the track A response group was overrepresented by the higher performing students. The track B response group scored somewhat lower on math ability than the non-response group, $t = 2.38$, $df\ 532$, $p < 0.05$ ($d = -0.55$), but other differences between the two groups were not significant ($p < 0.05$).

Descriptive results

Table 2 presents the correlations among the variables included in the study.

The highest correlations were established among the students' examination grades (between 0.45 and 0.60). There were moderate associations between the math and reading ability scores (0.40 in track A and 0.32 in track B) and small to moderate relations between both ability scores and the students' examination grades (correlations between 0.09 and 0.30). Furthermore, boys outperformed girls on math ability and physics grades while girls outperformed boys on reading ability. Finally,

TABLE 1

Descriptive statistics response and non-response group (means and standard deviations)

	Track A		Track B	
	Response	Non-response	Response	Non-response
Math ability	0.84 (0.86)	0.69 (1.12)	0.05 (0.83)	0.51 (0.86)
Reading ability	0.69 (0.92)	0.18 (0.85)	-0.28 (0.84)	-0.16 (0.79)
Examination grades ^a				
Advanced mathematics	6.6 (1.3)	6.3 (1.4)	6.3 (1.1)	6.2 (1.1)
Physics	6.8 (1.2)	6.6 (1.4)	6.6 (1.0)	6.5 (0.9)
Chemistry	6.6 (1.2)	6.4 (1.3)	6.4 (1.0)	6.2 (0.9)
N	931	189 ^b	515	117 ^b

^aThe grades range from 1 (lowest grade) to 10 (highest grade) (a 6 is a pass)

^bN is the amount of students for which all the variables stated were missing

TABLE 2
Correlations between the variables (track A in the upper cells; track B in the lower cells)

I	2	3	4	5	6	7
Math ability	1	0.40***	0.24***	0.30***	0.33***	-0.14***
Reading ability	0.32***	1	0.20***	0.28***	0.30***	-0.11***
Advanced mathematics	0.26***	0.10*	1	0.60***	0.54***	-0.02
Physics	0.18***	0.09*	0.50***	1	0.60***	-0.14***
Chemistry	0.15***	0.11*	0.45***	0.52***	1	-0.04
Sex ^a	-0.14**	0.09*	-0.03	-0.20***	-0.10*	1
Ethnicity ^b	-0.06	-0.11*	-0.08	-0.08	-0.07	-0.02
						1

^aPoint-biserial correlations (boys coded 0, girls coded 1)
^bPoint-biserial correlations (native Dutch students coded 0, students from ethnic minority groups coded 1)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

as regards students' ethnic background, native Dutch students outperformed students from ethnic minority groups on reading ability.

The multivariate multilevel models

Tables 3 (track A) and 4 (track B) show the results of the multivariate multilevel models. Model 0 was the unconditional model without predictors at the student level. Model 1 contained the predictor math ability. Subsequently, we added students' reading ability in Model 2. Finally, in Model 3, we also included the students' sex, ethnicity, and all two-way interaction terms. The asterisks indicate that the *t*-test statistics for the within-group regressions were significant.

After discussing the random parts of the models, we will subsequently deal with the fixed parts in order to test our hypotheses.

As regards both tracks, the random parts show variance in the students' examination grades at the student level and the school level. For example, as regards track A, Model 1 for advanced mathematics indicates an intraclass correlation coefficient (ICC) of $19.12/(19.12 + 151.09) = 0.11$. This coefficient signifies that in this example 11 % of the variance in the track A students' examination grades on advanced mathematics was found at the school level (i.e. the fraction of total variability that was located at the school level). Tables 3 and 4 include all ICCs. The variance at the school level for track A varies from 10 % to 12 % across the school subjects. For track B, the school level variance varies from 4 % to 11 %.

Next, we investigated the explained variances at the student level. For track A, for instance, Model 3 explained 8.0 % of the variance in the students' grades on advanced mathematics at the student level. This outcome was calculated as follows. The student level variance was $19.12 + 151.09 = 170.21$ in Model 0 and $17.47 + 139.07 = 156.54$ in Model 3. Thus, the explained variance at the student level was $1 - 156.54/170.21 = 0.080$ (8.0 % reduction). These variances are included in the tables. They vary from 2.3 % to 14.7 %, which is considered quite low for achievement measures. The explained variances in the track A models were, at least for the physics and chemistry grades, evidently higher than those in the track B models. We will get back to these findings in the "CONCLUSIONS AND DISCUSSION" section of this paper.

As regards the fixed parts of the unconditional models, we observed that the intercepts differed slightly among the school subjects for both tracks, where the highest mean was established for physics, followed by chemistry and advanced mathematics. The intercept of 64.98 for

TABLE 3

Parameter estimates (and standard errors) of the unconditional models and the models containing explanatory variables, indicated separately for advanced mathematics, physics, and chemistry (Track A; $N=931$)

[illegible]

TABLE 3
(continued)

Ethnicity							1.03	1.83
Reading ability×sex							−0.38	0.86
Reading ability×ethnicity							−0.08	1.28
Math ability×sex							−0.32	0.90
Math ability×ethnicity							−0.42	1.34
Sex×ethnicity							−1.13	2.23
Random effects								
Student level variance	123.70	5.99	112.19	5.42	107.46	5.20	107.13	5.19
School level variance	16.43	4.77	12.25	3.83	12.87	3.88	12.47	3.81
Intraclass correlation coefficient	0.12		0.10		0.11		0.10	
Explained variance student level			11.2 %		14.1 %		14.7 %	

^aThe GPA_{science} scores are multiplied by 10 for ease of interpretation of the coefficients (thus, they range from 10 to 100)

^b−2*loglikelihood (IGLS Deviance) values for the models are 20921.94 (Model 0), 20805.05 (Model 1), 20760.57 (Model 2), and 20643.29 (Model 3)

* $p < 0.05$

advanced mathematics in Model 0 (track A) should be interpreted as the expected value of the examination grade for a random student in a randomly drawn school (i.e. an average of 6.5). This intercept differed slightly from the raw mean of 6.6 on advanced mathematics presented in Table 1. This also applied to the other intercepts in the unconditional models. The reason for these small differences is that, in the estimation of the models presented in Tables 3 and 4, we also included the various schools, which was not done in the calculation of the raw means (Snijders & Bosker, 1999).

Tables 3 and 4 indicate that, for both tracks, the students' math ability was significantly related to their examination grades (see Model 1). The higher the score on math ability, the higher the examination grades on these subjects. Examining the effect of reading ability on the students' examination grades, we saw a clear difference between the tracks. For track A, the *t*test statistics (e.g. $1.69/0.49 = 3.45$ in Model 2 for advanced mathematics) revealed that reading ability had a significant effect on the students' examination grades on advanced mathematics, physics, and chemistry. For track B, we did not find reading ability effects on the students' grades. Apparently, the track B final examination grades were not directly affected by the students' reading ability, whereas those for track A students were. The coefficients indicate that, when a track A

TABLE 4

Parameter estimates (and standard errors) of the unconditional models and the models containing explanatory variables, indicated separately for advanced mathematics, physics, and chemistry (track B; $N = 515$)

Advanced mathematics	Model 0		Model 1		Model 2		Model 3	
Fixed effects	β	SE	β	SE	β	SE	β	SE
Intercept ^{ab}	62.58*	0.61	62.37*	0.60	62.45*	0.63	62.26*	0.78
Math ability			3.61*	0.59	3.53*	0.62	3.79*	0.84
Reading ability					0.29	0.62	−0.70	0.79
Sex							0.64	1.12
Ethnicity							−1.08	2.55
Reading ability×sex							2.09	1.28
Reading ability×ethnicity							1.81	2.41
Math ability×sex							−1.08	1.28
Math ability×ethnicity							1.02	2.24
Sex×ethnicity							−0.77	3.73
Random effects								
Student level variance	124.98	8.17	115.89	7.58	115.85	7.58	114.63	7.52
School level variance	5.46	3.64	5.96	3.60	5.95	3.59	6.19	3.64
Intraclass correlation coefficient	0.04		0.05		0.05		0.05	
Explained variance student level			6.6 %		6.6 %		7.4 %	
Physics	Model 0		Model 1		Model 2		Model 3	
Fixed effects	β	SE	β	SE	β	SE	β	SE
Intercept ^{ab}	65.56*	0.60	65.42*	0.59	65.61*	0.62	67.26*	0.78
Math ability			2.34*	0.54	2.15*	0.56	1.26	0.75
Reading ability					0.64	0.57	0.73	0.70
Sex							−3.97*	0.99
Ethnicity							−0.32	2.26
Reading ability×sex							0.01	1.13
Reading ability×ethnicity							1.46	2.13
Math ability×sex							0.34	1.13
Math ability×ethnicity							3.24	1.98
Sex×ethnicity							−1.64	3.29
Random effects								
Student level variance	97.01	6.37	93.59	6.15	93.26	6.12	88.60	5.83
School level variance	7.87	3.70	7.61	3.57	7.78	3.61	7.37	3.43
Intraclass correlation coefficient	0.08		0.08		0.08		0.08	
Explained variance student level			3.5 %		3.7 %		8.5 %	
Chemistry	Model 0		Model 1		Model 2		Model 3	
Fixed effects	β	SE	β	SE	β	SE	β	SE
Intercept ^{ab}	63.55*	0.65	63.44*	0.64	63.67*	0.66	64.45*	0.78
Math ability			1.80*	0.54	1.57*	0.56	1.65*	0.76
Reading ability					0.77	0.56	0.62	0.71

TABLE 4
(continued)

Sex							-1.86	1.01
Ethnicity							-2.53	2.28
Reading ability × sex							0.87	1.14
Reading ability × ethnicity							-1.09	2.15
Math ability × sex							-0.88	1.14
Math ability × ethnicity							0.57	2.01
Sex × ethnicity							-0.24	3.32
Random effects								
Student level variance	93.01	6.13	91.18	6.01	90.85	5.99	89.67	5.92
School level variance	11.83	4.48	11.25	4.31	11.15	4.28	11.28	4.13
Intraclass correlation coefficient	0.11		0.11		0.11		0.11	
Explained variance student level			2.3 %		2.7 %		3.7 %	

^aThe GPA_{science} scores are multiplied by 10 for ease of interpretation of the coefficients (thus, they range from 10 to 100)

^b-2*^aloglikelihood (IGLS Deviance) values for the models are 11296.36 (Model 0), 11258.19 (Model 1), 11258.98 (Model 2), and 11172.72 (Model 3)

* $p < 0.05$

student's reading ability score increased by one unit, his/her examination grades increased by 2 to 3 decimal points.

In Model 3, we included the variables sex, ethnicity, and two-way interaction terms. For track A, we observed that sex was significantly related to the students' examination grade on advanced mathematics (girls outperformed boys) but not to their physics and chemistry grades. Furthermore, neither ethnicity nor the interaction terms were significant in any of the models in this track. As regards track B, we found that sex was significantly related to the students' grade on physics (boys outperformed girls) but not to their advanced mathematics and chemistry grades. Ethnicity was again not significantly associated with the students' grades, and the interaction terms were non-significant. Finally, no significant effect of the interaction between the students' math and reading ability scores was observed on the students' examination grades.

CONCLUSIONS AND DISCUSSION

In the present study, we investigated to what extent students' math and reading ability relate to their final examination grades on the three school subjects mathematics, physics, and chemistry. In the textbooks of the

courses, contextualized word problems are commonly used. We hypothesized a positive association between the reading ability of the participants in our study and their examination grades on advanced mathematics, physics, and chemistry when controlling for their math ability (Hypothesis 1). Moreover, we expected this relation to be stronger for students from ethnic minority groups than for the native Dutch students and also stronger for boys than for girls (Hypothesis 2).

The significant contribution of the pre-university students' reading ability to their examination grades on the abovementioned subjects confirmed our expectation about the relationship between reading ability and achievement in advanced mathematics, physics, and chemistry (Hypothesis 1). This result is in line with previous research studies (Ní Ríordáin & O'Donoghue, 2009). The largest effects of reading ability were observed for physics and chemistry in track A, which suggests that the examinations in physics and chemistry are more 'language-oriented' than the advanced mathematics tests. As regards these three subjects, we did not find a significant relationship between reading ability and examination grades for the track B students. Hypothesis 1 has therefore not been confirmed for students in the senior general secondary education. It has to be mentioned that the FSEs differed between the track A and track B students, which could explain the anomalies between the tracks. However, large differences among the test items regarding the amount of word problems or overall reading requirements are unlikely. In general, the national examinations for these two tracks only differ in their levels of complexity and sometimes in their contents. A more plausible explanation is that, in addition to the track pursued, there were also other characteristics which caused the differences between the student samples. From the non-response analyses, it was concluded that the slightly better-performing track A students were overrepresented. Moreover, among the track A students, it was more common to take the FSE in advanced mathematics, physics, and chemistry than among the track B students (in the overall VOCL'99 sample 44 % versus 24 % of the track B students). Therefore, the track B students who did take their FSE in mathematics, physics, and chemistry were presumably those who had specific talents and/or interests in these courses, whereas for the track A students this combination of courses was generally considered as 'normal' (see also Korpershoek et al., 2011). This explanation is strengthened by the descriptive results presented in Table 1. The standard deviations of the examination grades of the track B students were somewhat lower than those of the track A students. Henceforth, we suggest a replication of the current study for a further exploration of the differences between the two tracks.

Additionally, based on our findings, we had to reject Hypothesis 2. Neither the main effects of ethnicity nor the two-way interaction effects of ethnicity and the ability measures on the students' examination grades were significant in the models. Furthermore, although two main effects of sex on the students' examination grades were significant (for track A on mathematics; for track B on physics), no effects of interactions between the ability measures and sex occurred. Therefore, we had to reject Hypothesis 2 that stated that the relation between students' reading ability and their examinations grades on advanced mathematics, physics, and chemistry would be stronger for students from ethnic minority groups than for native Dutch students and also stronger for boys than for girls. Presumably, students with relatively low reading ability had not entered track A and B in the first place, including students from ethnic minority groups with relatively low reading ability. A longitudinal study in which students' reading ability is measured at the end of primary education as well as each year in secondary education could reveal at which point (ethnic minority) students with low reading ability might become disadvantaged as compared with students with high reading ability.

A remarkable result we established is the fairly low level of explained variance in the examination grades in the models (it varied from 2 % to 15 %). This finding suggests that other explanatory variables such as subject-matter knowledge in advanced mathematics, physics, and chemistry and general cognitive skills (e.g. short term memory, Kintsch & Greeno, 1985) should ideally be included in the explanatory models to better understand the relative importance of math and reading ability in explaining students' academic achievement in these subjects. In this study, data measuring such variables were unfortunately unavailable. Furthermore, the relationships were plausibly not that strong because our sample generally consisted of relatively successful students (i.e. a restriction of range in reading and math ability). The students who had repeated a school year, dropped out of school, or had not taken their FSE in the 'difficult' school subjects, such as advanced mathematics, physics, and chemistry, were not included in the sample. Particularly in track B, it is uncommon to pursue science-oriented courses throughout secondary education. A third explanation is that the variance in the examination grades was relatively small, which suggests a restriction of the range of criterion variables. Although final examination grades theoretically range from 1 to 10, grades below 4 and above 9 are uncommon. This fact narrows down the variance in the examination grades that can be explained by the predictors. Finally, students who were at risk of failing their final examinations were presumably more motivated and spent more

hours preparing for their examinations than those who were likely to pass them. The former students must have been aware of the fact that they needed high grades in order to compensate for a low score already obtained on another course. In The Netherlands, it is in some cases allowed to compensate a low grade on one course with a high grade on another subject. Another notable finding is that across the models presented the ICCs varied, particularly for track B. For track B, the ICCs varied from 4–5 % in the mathematics models, and they were respectively 8 % and 11 % in the physics and chemistry models (for track A, all ICCs varied from 10–12 %). It seems that as regards the examination grades on advanced mathematics, it did not matter that much which school was attended by the students. For the grades on the other subjects and for track A specifically, the school attended seemed more important. Alternatively, the low percentages explained variance in our models could be the result of the way we measured students' math and reading ability and the procedure we followed to calculate the ability scores for all students, also for the ones who completed only two out of three tests. Although the three predictor tests that were used to calculate (weighted) ability scores shared much common variance (65 % in the math ability score, 57 % in the reading ability score, respectively), the correlations among the three predictor tests were moderate for mathematics and low to moderate for reading. Hence, the predictive validity of our constructed ability measures may have been limited due to measurement errors in measuring the underlying constructs (particularly for reading ability). Although the final school examinations generally aim at measuring a broad range of subject-specific abilities among secondary school students, our indicators of students' ability may not have covered all ability aspects that were assessed in the examination. This appears to be case particularly for track B students, since we could not detect any significant effects of reading ability on track B students' final examination grades on advanced mathematics, physics, and chemistry. Following our results and the addressed measurement limitations, an in-depth analysis of the reading ability demand of the final examinations in each educational track may shed some light on this matter (see e.g. Lucangeli, Tressoldi & Cendron, 1998, who studied students' word problem solving by investigating students' cognitive abilities [i.e. text comprehension] and metacognitive abilities [e.g. planning the solution] as predictors).

To sum up, our findings suggest that students with a high score on reading ability benefit from this skill when taking their FSEs in advanced mathematics, physics, and chemistry, at least the pre-university students do. Making sense of mathematical word problems evidently requires a

certain level of text comprehension (see also Verschaffel, Greer & De Corte, 2000). To construct meaning from the story or real-life scenario presented in mathematical word problems, students obviously need a certain level of reading ability (Adams, 2003; Graesser, Millis & Zwaan, 1997). Based on our findings, it may be justifiable to argue that reading comprehension should be an integrated part of mathematics, physics, and chemistry courses (see also OECD, 2009). In the literature, the impact of reading ability on students' final examination grades on science-related school subjects has not received much attention in previous studies, since the samples used in most studies that addressed the reading-math/science relation were primary education samples (e.g. Kintsch & Greeno, 1985; Lee et al., 2005; Powell et al., 2009; Stern, 1993; Vilenius-Tuohimaa et al., 2008; Walker et al., 2008). In the present study, we therefore sought to contribute to the literature by testing the predictive power of students' math ability and reading ability in explaining their examination grades on advanced mathematics, physics, and chemistry in secondary education. Particularly the fact that we used not only students' examination grades on advanced mathematics but also on physics and chemistry as outcome measures is new in the field. A notable result is that students' math ability had a significant impact on all these three outcome measures, indicating that a certain level of mathematical understanding is required for performing well at physics and chemistry exams. Moreover, to our knowledge, most papers addressed the reading-math/science relation by using one single indicator of students' ability in a particular domain (e.g. a mathematics test and a reading test), whereas our results confirm that different performance tests within a certain domain can measure different aspects of a single underlying construct. In our study, we constructed two domain-specific ability measures based on a range of different ability tests and an intelligence test to address this issue. Despite the fact that the predictor power of our measures was limited, our approach to use weighted test scores seemed suitable for integrating different aspects of an underlying ability construct in one single predictor.

There are a number of limitations in addition to the stated measurement issues to consider when generalizing the present findings. First, following the Dutch curriculum the students in our sample were obliged to take their FSEs in the fixed combination of advanced mathematics, physics, and chemistry. This might be different in other countries. Therefore, generalizing the results to international student samples should be made with some caution. Furthermore, we did not look into the reading skills required for the test items included in the FSEs (e.g. Walker et al., 2008). Following our results, it would be helpful to analyze the level of text

comprehension required for the test items used in the mathematics, physics, and chemistry examinations. Furthermore, we were unable to obtain individual item scores from the VOCL'99 database, which limited our possibilities to narrow down the type of questions that caused the most problems. Finally, as the design of the current research was non-experimental, we underline that any suggestions for educational practice based on this sole study are premature. Future studies can, however, build on the findings presented here. We suggest developing (quasi-) experimental studies with different conditions aimed at enhancing students' mathematics, physics, and chemistry understanding. Relevant experimental conditions include teachers paying particular attention to the required math, physics, and chemistry vocabulary in their lessons and teachers providing additional guidance for students with reading difficulties (e.g. Stoddard et al., 2002). Furthermore, a comparative study evaluating student achievement on two tests that differ in the required comprehensive reading skills yet are similar in difficulty and content could narrow down the type of word problems that cause most problems for students with reading difficulties (e.g. Shaw et al., 2010; Turkan & Liu, 2012).

Mathematical word problems are tools for practicing in dealing with all kinds of situations in everyday life (i.e. mathematical modeling). Their general aim is to train students in thinking creatively and help them grasp new mathematical concepts and skills (Verschaffel et al., 2000). Although several critics have questioned the extent to which word problems actually achieve these aims (e.g. Gerofsky, 1996), students in secondary education are confronted with them on a daily basis in the lessons they receive. Therefore, further examination of the relationship investigated in this study is certainly needed to increase our understanding of the significance of reading ability for educational attainment in secondary education.

NOTES

¹ In the literature, different words for reading ability are used, such as language proficiency, language skills, language literacy, reading comprehension, reading skills, etc. For ease of interpretation, we will mostly use the term *reading ability* throughout this manuscript.

² Although biology textbooks also include contextualized word problems, biology was not included in the present study because only part of our student sample had taken the national examinations in this subject.

REFERENCES

- Adams, T. L. (2003). Reading mathematics: More than words can say. *Reading Teacher*, 56, 786–795.
- Baumert, J., Lüdtke, O., Trautwein, U. & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review*, 4, 165–176.
- Beal, C. R., Adams, N. M. & Cohen, P. R. (2010). Reading proficiency and mathematics problem solving by high school English language learners. *Urban Education*, 45, 58–75.
- Bell, R. L., Matkins, J. J. & Gansneder, B. M. (2011). Impacts of contextual and explicit instruction on preservice elementary teachers' understandings of the nature of science. *Journal of Research in Science Teaching*, 48, 414–436.
- Belloccchi, A. & Ritchie, S. M. (2011). Investigating and theorizing discourse during analogy writing in chemistry. *Journal of Research in Science Teaching*, 48, 771–792.
- Borasj, R. & Siegel, M. (1998). Using transactional reading strategies to support sense-making. *Journal for Research in Mathematics Education*, 29, 275–305.
- Brown, B. A. & Ryoo, K. (2008). Teaching science as a language: A “content-first” approach to science teaching. *Journal of Research in Science Teaching*, 45, 529–553.
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Carter, T. & Dean, E. (2006). Mathematics intervention for grades 5–11: Teaching mathematics, reading, or both? *Reading Psychology*, 27, 127–146.
- Chapman, O. (2006). Classroom practices for context of mathematics word problems. *Educational Studies in Mathematics*, 62, 211–230.
- Cronbach, L. J., Schönemann, P. & McKie, D. (1965). Alpha coefficients for stratified parallel tests. *Educational and Psychological Measurement*, 25, 291–312.
- Evers, A., van Vliet-Mulder, J. C. & Groot, C. J. (2000). *Documentatie van tests en testresearch in Nederland [Documentation of tests and test research in The Netherlands]*. Assen, The Netherlands: Van Gorcum.
- Fagan, D. (1997). Reading in advanced level physics. *Physics Education*, 32, 383–386.
- Flick, L. B. & Lederman, N. G. (2002). The value of teaching reading in the context of science and mathematics. *School Science & Mathematics*, 102, 105–106.
- Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J., Hamlett, C. L. & Lambert, W. (2008). Problem-solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology*, 100, 30–47.
- Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics*, 16, 36–45.
- Graesser, A. C., Millis, K. K. & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163–189.
- Grimm, K. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, 33, 410–426.
- Helwig, R. & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research*, 93, 113–125.
- Jacobson, M. D. (1965). Reading difficulty of physics and chemistry textbooks. *Educational and Psychological Measurement*, 25, 449–457.

- Kamata, A., Turhan, A. & Darandari, E. (2003). *Estimating reliability for multidimensional composite scale scores*. Paper presented at the annual meeting of American Educational Research Association, Chicago, IL.
- Kintsch, W. & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92, 109–129.
- Koch, A. & Eckstein, S. G. (1995). Skills needed for reading comprehension of physics texts and their relation to problem solving ability. *Journal of Research in Science Teaching*, 32, 613–628.
- Koch, A. & Gunstone, R. (2001). Training in metacognition and comprehension of physics texts. *Science Education*, 85, 758–768.
- Korpershoek, H., Kuyper, H. & van der Werf, M. P. C. (2006). *HAVO-5 en VWO-5 en de tweede fase; De bovenbouwstudie van VOCL '99 [HAVO-5 and VWO-5 and the second phase; the study in the upper grades of VOCL '99]*. Groningen, The Netherlands: GION.
- Korpershoek, H., Kuyper, H., van der Werf, M. P. C. & Bosker, R. J. (2011). Who succeeds in advanced mathematics and science courses? *British Educational Research Journal*, 37, 357–380.
- Kurth, L. A., Kidd, R., Gardner, R. & Smith, E. L. (2002). Student use of narrative and paradigmatic forms of talk in elementary science conversations. *Journal of Research in Science Teaching*, 39, 793–818.
- Kyttälä, M. (2008). Visuospatial working memory in adolescents with poor performance in mathematics: Variation depending on reading skills. *Educational Psychology*, 28, 290–307.
- Lantz-Andersson, A., Linderöth, J. & Säljö, R. (2009). What's the problem? Meaning making and learning to do mathematical word problems in the context of digital tools. *Instructional Science*, 37, 325–343.
- Lee, O. (2004). Teacher change in beliefs and practices in science and literacy instruction with English language learners. *Journal of Research in Science Teaching*, 41, 65–93.
- Lee, O., Deaktor, R. A., Hart, J. E., Cuevas, P. & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, 42, 857–887.
- Lucangeli, D., Tressoldi, P. E. & Cendron, M. (1998). Cognitive and metacognitive abilities involved in the solution of mathematical word problems: Validation of a comprehensive model. *Contemporary Educational Psychology*, 23, 257–275.
- Mullis, I. V. S., Martin, M. O. & Foy, P. (2008). *TIMSS 2007 International mathematics report. Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Nathan, M. J., Kintsch, W. & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9, 329–389.
- Ní Riordáin, M. & O'Donoghue, J. (2009). The relationship between performance on mathematical word problems and language proficiency for students learning through the medium of Irish. *Educational Studies in Mathematics*, 71, 43–64.
- Ogborn, J. M. (1996). *Explaining science in the classroom*. Buckingham: Open University Press.
- Organisation for Economic Co-operation & Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: Author.
- Organisation for Economic Co-operation & Development. (2007). *PISA 2006 Science competencies for tomorrow's world*. Paris, France: Author.

- Organisation for Economic Co-operation & Development. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris, France: Author.
- Ozuru, Y., Dempsey, K. & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19, 228–242.
- Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T. & Fletcher, J. M. (2009). Do word-problem features differentially affect problem difficulty as a function of students' mathematics difficulty with and without reading difficulty? *Journal of Learning Disabilities*, 42, 99–110.
- Reikerås, E. K. L. (2006). Performance in solving arithmetic problems: A comparison of children with different levels of achievement in mathematics and reading. *European Journal of Special Needs Education*, 21, 233–250.
- Reusser, K. & Stebler, R. (1997). Every word problem has a solution—The social rationality of mathematical modeling in schools. *Learning and Instruction*, 7, 309–327.
- Shaw, J. M., Bunch, G. C. & Geaney, E. R. (2010). Analyzing language demands facing English learners on science performance assessments: The Sald framework. *Journal of Research in Science Teaching*, 47, 909–928.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London, UK: SAGE Publications.
- Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology*, 85, 7–23.
- Stoddard, T., Pinal, A., Latzke, M. & Canaday, D. (2002). Integrating inquiry science and language development for English language learners. *Journal of Research in Science Teaching*, 39, 664–687.
- Turkan, S. & Liu, O. L. (2012). Differential performance by English Language Learners on an inquiry-based science assessment. *International Journal of Science Education*, 34, 2343–2369.
- Van Berkel, K. (1999). *Steekproef voor VOCL '99 [Sample for VOCL '99]*. Heerlen: Statistics Netherlands (CBS).
- Van Dijk, H. & Tellegen, P. J. (1994). *Handleiding, testboekje, instructieboekje GIVO, Groninger intelligentietest voor voortgezet onderwijs. [Manual, test book, instruction book GIVO, Groninger intelligence test for secondary education]*. Lisse: Swets & Zeitlinger.
- Verschaffel, L., Greer, B. & de Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.
- Vilenius-Tuohimaa, P. M., Aunola, K. & Nurmi, J. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, 28, 426–443.
- Walker, C. M., Zhang, B. & Surber, J. (2008). Using a multidimensional differential item functioning framework to determine if reading ability affects student performance in mathematics. *Applied Measurement in Education*, 21, 162–181.
- Yore, L. D. (1991). Secondary science teachers' attitudes toward and beliefs about science reading and science textbooks. *Journal of Research in Science Teaching*, 28, 55–72.