



Self-beliefs mediate math performance between primary and lower secondary school: A large-scale longitudinal cohort study

Helen C. Reed^a, Paul A. Kirschner^b & Jelle Jolles^a

^a VU University Amsterdam, The Netherlands

^b Welten Institute, Open University of the Netherlands

Article received 10 December / revised 19 February / accepted 5 April / available online 21 April

Abstract

It is often argued that enhancement of self-beliefs should be one of the key goals of education. However, very little is known about the relation between self-beliefs and performance when students move from primary to secondary school in highly differentiated educational systems with early tracking. This large-scale longitudinal cohort study examines the extent to which academic self-efficacy (i.e., how confident students are that they will be able to master their schoolwork) and math self-concept (i.e., students' perceived math competence) mediate the relation between math performance at the end of primary school (Grade 6) and the end of lower secondary school (Grade 9) in such a system. The study involved 843 typically-developing students in the Netherlands. Self-efficacy and math self-concept were measured with self-report questionnaires. Math performance was measured with nationally validated tests. The relation between math performance in Grade 6 and in Grade 9 was uniquely mediated by both self-efficacy in Grade 6 and math self-concept in Grade 9, but in opposing directions. Math self-concept was the most influential mediator, explaining nearly a quarter of the total effect of Grade 6 math performance on Grade 9 math performance. Unexpectedly, high self-efficacy in Grade 6 was negatively related to Grade 9 math performance, particularly for girls and high-track students. These findings suggest that self-efficacy may not necessarily be a protective factor in highly differentiated early tracking educational systems and may need to be actively managed when students move to secondary school.

Keywords: self-beliefs; self-efficacy; math self-concept; math performance; school transition; educational tracking

Corresponding author: Helen C. Reed, Department of Educational Neuroscience and LEARN! research institute, Faculty of Psychology and Education, VU University Amsterdam, Van der Boerhorststraat 1, 1081 BT Amsterdam, The Netherlands; E-mail: HC.Reed@vu.nl Doi <http://dx.doi.org/10.14786/flr.v3i1.139>



1. Introduction

Most students hold beliefs about their own capabilities and competence in accomplishing academic tasks. Do these so-called *self-beliefs* affect the relation between students' performance at the end of primary school and the end of lower secondary school? This question is especially relevant in systems that make use of early educational tracking to stratify students according to scholastic ability. Tracking is based on the premise that homogeneous classes allow curriculum and instruction to be directed towards the common needs of groups of similar ability and that this leads to maximum learning for all (Chmielewski, Dumont, & Trautwein, 2013; Hanushek & Wößmann, 2006). Tracking is considered highly differentiated when students are stratified into different schools or educational programs with little or no contact between them.

In educational systems with early tracking (e.g., the Netherlands (see Box 1), Germany, Belgium (Flanders), Singapore), track placement in lower secondary school depends to a large extent on performance at the end of primary school or even earlier. An important assumption is that there is a substantial degree of stability in performance between the end of primary school and lower secondary school. If this were not the case, then discrepancies between track placement and students' actual performance would soon render these systems ineffectual.

The stability of this relation could, however, be affected by student variables that depress or elevate performance in secondary school relative to expectations at the moment of track assignment. Of particular concern are students whose performance in secondary school falls below expectation. Students who fail in their designated track are often retained or drop down to a lower track, which is reported to be detrimental to student outcomes (Brophy, 2006; Jacob & Lefgren, 2009; OECD, 2012). It may be possible to prevent this happening when more is known about student variables that affect the stability of the relation between performance in primary and secondary school. Within this context, the present study investigates the extent to which the relation between math performance at the end of primary school (i.e., Grade 6) and the end of lower secondary school (i.e., Grade 9) in a highly differentiated early tracking educational system is mediated by student self-beliefs relating to their academic functioning at school.

Box 1: Educational tracking in the Netherlands

In the Netherlands, educational tracking is implemented early in secondary school. Track placement depends largely on performance at the end of primary school and is based on school grades and/or the results of a school placement test, as well as study skills, concentration, motivation, application, etcetera. Once track placement has been determined - sometimes after an initial orienting period - there is little academic contact (e.g., shared classes) between tracks. There are three main tracks: *pre-university* (preparing the most able students for university; 6 years duration; around 20% of students), *higher general secondary* (preparation for professional higher education; 5 years duration; 20%), and *pre-vocational* (theory-oriented or practice-oriented preparation for vocational education; 4 years duration; 55%). In addition, around 5% of students are in special needs education or receive training in low-level practical skills for entry to the workforce.

1.1 Self-beliefs in school settings

A large body of research indicates that positive self-beliefs are strongly related to higher academic performance, as we review presently. It is therefore worrying that many students experience a decline in self-beliefs between primary and secondary school (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Liu, Wang, & Parkins, 2005), especially in the domain of mathematics. For example, the most recent cycle of the Trends in International Mathematics and Science Study reported that only just over a tenth of 8th graders are confident in their mathematics ability compared to a third of 4th graders (Mullis, Martin, Foy, & Arora, 2012).



The causes of this decline appear to be manifold. When students move from primary to secondary school, they are confronted with many factors (e.g., different learning and assessment goals, demands and conditions; relationships with peers and teachers; biological and neurological changes of adolescence) that can affect their beliefs about their ability to do well in the new school environment (Cauley & Jovanovich, 2006; Fenzel, 2000; Sakiz, Pape, & Hoy, 2012; Schunk & Meece, 2006; Urdan & Schoenfelder, 2006). The early years of secondary school occur at a crucial developmental period in early to mid-adolescence. On the one hand, adolescents have greater need for autonomy, feelings of competence, social connectedness, and positive relations with peers and adults; while at the same time they have heightened sensitivity to social comparisons, peer influence and emotional support or the lack thereof (Cauley & Jovanovich, 2006; Osterman, 2000; Sakiz et al., 2012; Schunk & Meece, 2006). On the other hand, secondary schools are more anonymous and more regimented than primary schools, there is stronger emphasis on testing and grades, teachers are perceived as more controlling and distant and less supportive and fair, and schoolwork is more plentiful and more demanding (Cauley & Jovanovich, 2006; Osterman, 2000; Sakiz et al., 2012).

From a developmental perspective, demands made on adolescent learners could diverge from their neurocognitive capacities to meet them. For instance, many academic areas (e.g., science and mathematics, language learning and problem-solving) require higher-order thinking skills that depend on neural networks which show considerable individual variability in maturation during adolescence (e.g., Crone et al., 2009; Dumontheil, Houlton, Christoff, & Blakemore, 2010). If students are required to think in ways that exceed their developmental capabilities, frustration, disillusionment, and decreased feelings of competence can result (Cauley & Jovanovich, 2006). Furthermore, secondary school students are often expected to regulate their own learning at a time when their behavioural control is compromised by a heightened sensitivity to motivational cues (Somerville & Casey, 2010). In short, a mismatch between students' developmental needs and capacities and the secondary school environment can lead to reduced motivation, engagement, interest in school, and beliefs about their ability to succeed (Cauley & Jovanovich, 2006; Sakiz et al., 2012; Schunk & Meece, 2006; Urdan & Schoenfelder, 2006).

The present study focuses on two of the most influential and widely studied types of self-beliefs, namely *self-efficacy* and *self-concept*, both of which arise from the perception and appraisal of oneself in relation to prior experience (Huang, 2011; Marsh & Martin, 2011; Valentine, DuBois, & Cooper, 2004). Within the school context, *self-efficacy* refers to what individuals expect and believe they will be able to accomplish in academic tasks with whatever abilities and skills they may have (Bandura, 1997; Bong & Skaalvik, 2003; Schunk & Meece, 2006). It is typically measured by asking individuals to judge how confident they are that they will be able to master their schoolwork or perform representative tasks. *Self-concept* represents an individual's evaluation of their actual functioning or competence in general or in a specific domain (Bong & Skaalvik, 2003; Marsh & Martin, 2011). It is typically measured by asking individuals to indicate the extent to which they endorse statements as "I am good at (a particular subject area)". Thus, the conviction that one will be able to pass a test if one studies for it is a self-efficacy judgment, while the belief that one is not good at math is a self-concept judgment.

Self-efficacy and self-concept are not always clearly distinguished in the literature. Nonetheless, a comprehensive review by Bong and Skaalvik (2003) identified important differences between the constructs. These include the extent to which they are influenced by goals and designated standards, social norms, and/or internal comparisons (e.g., comparing one's own performance in different domains or across time); whether they are oriented to the future (i.e., what one believes one could achieve) or to the past (i.e., what one has actually achieved); and whether they are changeable or stable across time. In these terms, self-efficacy is argued to be heavily goal-referenced, somewhat normatively referenced, future-oriented and temporally changeable. By comparison, self-concept is both normatively and ipsatively referenced, past-oriented and more stable across time.

Despite these differences, self-efficacy and self-concept share important similarities. For one, they are both shaped by individuals' prior experiences and performance (Bong & Skaalvik, 2003; Möller, Pohlmann, Köller, & Marsh, 2009; Möller, Retelsdorf, Köller, & Marsh, 2011; Schunk & Meece, 2006). For



example, self-efficacy is strengthened by successful experiences and undermined by repeated failures, while self-concept in particular academic areas (e.g., mathematics, languages, science) is influenced by students' achievement in these areas over time (Möller et al., 2009; Möller et al., 2011; Skaalvik, & Skaalvik, 2002). Another important antecedent of both constructs is the appraisal of significant others such as parents and teachers, which can influence and/or reinforce individuals' views of themselves (Bong & Skaalvik, 2003). Thus, for instance, when teachers express that a student will succeed and is good at certain things, this can contribute to the student's own expectation of success and positive appraisal of his/her abilities.

Self-efficacy and self-concept are also both influenced by comparisons in relation to personally relevant external frames of reference, notably similar peers (Möller et al., 2009; Möller et al., 2011; Skaalvik, & Skaalvik, 2002). Thus, students' self-efficacy beliefs may be influenced by the performance of similar classmates on particular tasks: when classmates are successful, students may become more confident that they too will succeed on the tasks in question, and when classmates are unsuccessful, students may become less confident of success. Similarly, students' self-concept in a particular subject area is shaped through comparing their own achievements to those of their classmates: if their math achievement is higher than that of their classmates, math self-concept is generally also higher. In the initial years of secondary school, peer comparisons are particularly influential because students are unfamiliar with many of the tasks and learning environments and have few sources of information other than their friends with which to gauge their own experiences (Schunk & Meece, 2006).

These issues are complicated in highly differentiated early tracking educational systems where students move from heterogeneous primary school classrooms into ability-homogeneous tracks in secondary school. Under these circumstances, peer comparisons are affected by the so-called 'Big-Fish-Little-Pond' effect (Marsh, 1991; Marsh & Hau, 2003). This refers to the phenomenon that performance of higher ability students in mixed ability groups is higher than most of their classmates, which elevates self-judgments in comparison to others. However, their performance may be only average or below-average in groups whose performance standards are set by high ability students; self-judgments are then likely to be lower. The reverse is true for lower ability students. Thus, the change in reference peer group after the move from primary to secondary school in early tracking systems is likely - over time - to depress self-beliefs in higher tracks and increase them in lower tracks. This is particularly so where there is little academic contact between students in different tracks (as in the Netherlands), so that within-track - as opposed to across-track - comparisons become dominant (Chmielewski et al., 2013; Liu et al., 2005). Investigating self-beliefs within a system of educational tracking therefore requires careful consideration of the effects of changes in reference group.

1.2 Self-beliefs and math performance

Previous research has demonstrated strong relationships between self-beliefs and academic performance generally (Caprara, Vecchione, Alessandri, Gerbino, & Barbaranelli, 2011; Huang, 2011; Marsh & Martin, 2011; OECD, 2013; Schunk & Meece, 2006; Valentine et al., 2004) as well as between math-related self-beliefs and math performance specifically (Chiu & Klassen, 2010; Ferla, Valcke, & Cai, 2009; Ireson & Hallam, 2009; Möller et al., 2009; Möller et al., 2011; Skaalvik & Skaalvik, 2006; Steinmayr & Spinath, 2009; Valentine et al., 2004). Self-beliefs and performance are more strongly related when measured at the same level of specificity (Bong & Skaalvik, 2003; Valentine et al., 2004). Thus, general self-beliefs - such as the belief that one will be able to master one's schoolwork - are less strongly related to math performance than the specific belief that one is good (or not good) at math.

Importantly, and as a point of departure for the present research, reciprocal effects between math performance and self-beliefs have been demonstrated in longitudinal studies (Marsh & Martin, 2011; Möller et al., 2011; Pajares & Schunk, 2001). These studies indicate that: (a) math performance at an earlier time point affects math performance at a later time point; (b) math performance influences students' self-beliefs; and (c) students' self-beliefs affect math performance. While these studies have established that self-beliefs



mediate the relation between math performance at successive time points - presumably by means of mutual reinforcement - there is currently little research that examines these effects in highly differentiated early educational tracking systems spanning the period bridging primary and secondary school. As noted, the change in reference group when students move from heterogeneous primary school classrooms to homogeneous secondary school classrooms can profoundly affect students' self-beliefs through the mechanism of peer comparison. Thus, research still needs to resolve the role of self-beliefs in this situation.

Finally, it is possible that the relation between self-beliefs and math performance could be moderated by sex (Valentine et al., 2004). Boys and girls differ in self-beliefs in several academic areas, including mathematics (Herbert & Stipek, 2005; Ireson & Hallam, 2009; Jacobs et al., 2002; Preckel, Goetz, Pekrun, & Kleine, 2008; Schunk & Meece, 2006). Moreover, girls report lower self-belief in their math competence than boys, even when performance levels are equal (Else-Quest, Hyde, & Linn, 2010; OECD, 2013). Thus, self-beliefs could have different effects on math outcomes for boys and girls.

1.3 The present study

The present study addresses these issues by examining the extent to which self-efficacy (i.e., how confident students are that they will be able to master their schoolwork) and math self-concept (i.e., students' perceived math competence) mediate the relation between math performance at the end of primary school (i.e., Grade 6) and the end of lower secondary school (i.e., Grade 9) in a highly differentiated early tracking system. This is investigated in a multiple mediator model reflecting the reciprocal effects identified above and including self-belief measures in Grade 6 and Grade 9. Furthermore, the study examines whether these relations are moderated by educational track and/or sex.

The study draws on a large sample of typically-developing students who participated in a nationally representative, longitudinal cohort study in the Netherlands. Next to the longitudinal design, a strength of the study is that math performance was measured with validated, standardised national tests rather than school grades, which are known to suffer from variability in assessment and grading practices (Bowers, 2011). The measures used here can therefore be considered a more reliable proxy for math performance. Furthermore, performance was standardised within relevant reference peer groups. This is a crucial point, given the importance of these frames of reference in shaping self-beliefs.

The large-scale longitudinal design combined with the use of validated measures allows strong inferences to be drawn about the relations of interest within the context of highly differentiated early educational tracking. The results could therefore be of considerable value in identifying factors that could affect students' ability to maintain the levels of secondary school performance that are expected in their designated track.

2. Methods

This study comprises secondary analysis of data from the first and second cohort measurements of the COOL⁵⁻¹⁸ study (Cohort Research on Educational Careers), a large-scale, nationally representative, longitudinal cohort study into the determinants of the cognitive and social-emotional development of children and adolescents in the Netherlands¹. The COOL⁵⁻¹⁸ datasets are available for third-party use, as in the present study. The first cohort measurement included $N = 11,609$ Grade 6 students from 550 primary schools. The second measurement included $N = 21,384$ Grade 9 students from 151 secondary schools. A total of $N = 2,646$ students from 355 primary schools and 143 secondary schools participated in the first measurement when in Grade 6 and in the second measurement when in Grade 9. Participants took several cognitive tests at each measurement, including a math test. They also completed self-report questionnaires



that included scales from externally validated questionnaires on topics including self-efficacy and school functioning. Parents/caregivers completed a demographic questionnaire and schools provided administrative data (e.g., age, sex, educational track). The following paragraphs describe the participants, instruments and data relevant to the present study.

2.1 Participants

Individuals were selected when they had participated in the COOL⁵⁻¹⁸ study in both Grade 6 and Grade 9, when they had Dutch nationality and when complete data were available for sex, educational track, both math tests (i.e., in Grade 6 and Grade 9), and the hypothesised mediators (i.e., self-efficacy and math self-concept). In addition, students had to be aged between 14.5 and 15.5 years at Grade 9 measurement. An age-restricted window was chosen in order to have a relatively homogeneous sample of typically-developing students. Accelerated and delayed students were excluded, as these students differ from their classmates in several respects relating to self-beliefs that could confound the results. For example, delayed secondary school students have significantly lower self-beliefs about their ability to do well in school (Martin, 2011), while accelerated students in Dutch lower secondary school have more positive self-beliefs about their school abilities and their math ability in particular (Hoogeveen, Van Hell, & Verhoeven, 2009). Of the $N = 969$ students for whom the required data were available, 78 (8%) delayed students and 35 (3.6%) accelerated students were excluded. Another 13 (1.3%) students were excluded as age was unknown.

The final sample comprised $N = 843$ students (47% male ($N = 394$); $M_{age} = 14.9$ years, $SD_{age} = 0.3$). Of these, $N = 329$ (39%) were in a 'low' track (i.e., pre-vocational education), $N = 235$ (28%) were in a 'medium' track (i.e., higher general secondary education) and $N = 279$ (33%) were in a 'high' track (i.e., pre-university education). The students came from 188 primary schools and 101 secondary schools.

2.2 Grade 6 instruments and data

2.2.1 Math performance

Participants were administered a validated, standardised, norm-referenced math test for Grade 6 (M8, 2002 version) developed by the Dutch Central Institute for Educational Measurement. The test contained 107 items covering: (1) numbers and number relations; (2) arithmetic fact fluency; (3) mental arithmetic; (4) multiple operations; (5) fractions; (6) proportions; (7) percentages; (8) measurement; (9) geometry; (10) time. Raw test scores were converted to proficiency scores (range: 54-160) according to standard procedure. One case with an input error was excluded from analysis.

As indicated, students' self-beliefs are influenced by comparison of their own achievements relative to relevant reference peer groups. Thus, proficiency scores were standardised to denote individual performance relative to performance levels of these reference groups. In Grade 6 (before stratification), class or school can be considered a relevant reference group. In the COOL⁵⁻¹⁸ dataset, distribution of participants across classes was uneven, so scores were standardised per school within the whole Grade 6 sample. This approach effectively nests students within schools. The whole sample (i.e., before exclusion of participants on grounds of missing data, nationality or age) was used for standardisation to keep reference groups intact. The standardised scores were used as the Grade 6 math performance measure. The correlation between the standardised and unstandardised scores was high ($r = .82, p < .001$).

2.2.2 Self-efficacy

Self-efficacy was measured by the academic efficacy scale of the Patterns of Adaptive Learning Scales (PALS; Midgley et al., 2000; Urdan & Midgley, 2003) from the student questionnaire. This instrument has strong psychometric properties and strong predictive and concurrent validity for both primary and secondary school students (Anderman, Urdan, & Roeser, 2003) and is therefore highly suitable for the



present purpose. The self-efficacy scale contains six items (e.g., “*I’m certain I will be able to master the skills taught in school this year*” and “*I’m certain I could figure out how to do even the most difficult classwork*”), rated on a 5-point Likert-type scale with choice options ranging from ‘not at all true’ to ‘very true’. Items (in Dutch) were coded from 1 to 5, with higher scores indicating higher self-efficacy. Scale internal reliability was acceptable (Cronbach’s $\alpha = .78$). Self-efficacy was calculated as the average of the six items.

2.3 Grade 9 instruments and data

2.3.1 Math performance

Participants were administered a validated, norm-referenced math test developed by the Dutch Central Institute for Educational Measurement. Test items were drawn from an item-bank of 60 items and administered in three test versions comprising 30 multiple-choice items on arithmetic, proportions, geometry and mathematical relationships. An example (translated from the original Dutch) is:

A group of 5 men buys one lottery ticket between them every month. A group of 8 women does the same. There is one lottery draw per month. If a prize is won by one of the tickets, then the prize is shared out among the group members: among the 5 members of the men’s group and among the 8 members of the women’s group. In April, the ticket bought by the men’s group won a prize of €100,000 and the ticket bought by the women’s group won a prize of €200,000. Each man then received an amount of money and each woman received another amount of money. The amount that each man received was:

A $\frac{4}{5}$ times

B $\frac{5}{8}$ times

C $\frac{5}{4}$ times

D $\frac{8}{5}$ times

the amount that each woman received.

As not all participants were administered the same test version, their test scores would not be comparable under standard scoring procedures. Thus, items were analysed using the One-Parameter Logistic Model (OPLM; Verhelst & Glas, 1995) from Item Response Theory. When the OPLM holds for a collection of test items, a student’s skill level can be estimated from every subset of items - in this case, each test version. The OPLM was used to translate raw test scores to skill-scores that in turn were translated to bank-scores on a scale of 0 to 100% (Hambleton, Swaminathan, & Rogers, 1991). The bank-score indicates individual mastery level (e.g., a bank-score of 70 means that the student is expected to answer 70% of the total item-bank correctly) and is directly comparable across participants and test versions.

In Grade 9 in the Netherlands, relevant reference groups are class or the school/track combination within which classes are embedded. Again, distribution of participants across classes in the COOL⁵⁻¹⁸ dataset was uneven, so bank-scores were standardised per school and track within the whole Grade 9 sample to denote performance relative to this reference group. This approach nests students within school and educational track. The whole sample (i.e., before exclusion of participants on grounds of missing data, nationality or age) was used for standardisation to keep reference groups intact. The standardised scores were used as the Grade 9 math performance measure. The correlation between the standardised and unstandardised scores ($r = .56, p < .001$) was lower than that between the standardised and unstandardised Grade 6 measures. This is consistent with the fact that standardised Grade 9 scores were relative to scores of students of similar ability level (i.e., track) rather than being relative to scores of students of all ability levels, as in Grade 6.



2.3.2 Self-efficacy

Self-efficacy was measured by the academic efficacy scale of the Patterns of Adaptive Learning Scales from the student questionnaire, coded as described above. Scale internal reliability was acceptable (Cronbach’s $\alpha = .83$). Self-efficacy was calculated as the average of the scale items.

2.3.3 Math self-concept

Math self-concept was measured by the item: “*I am good at arithmetic and math*” (in Dutch) from the student questionnaire, with choice options ‘disagree’, ‘partly agree’ and ‘agree’. The item was coded from 1 to 3, with higher scores indicating a higher competence judgment. Single-item measures are frequently used in research on self-beliefs, for example by having participants indicate an anticipated exam grade (e.g., Vancouver & Kendall, 2006). A single omnibus measure can be as psychometrically sound and effective as multiple-item measurement scales in self-report questionnaires (Gardner, Cummings, Dunham, & Pierce, 1998; Robins, Hendin, & Trzesniewski, 2001) and can eliminate item redundancy and variance due to spurious correlations between highly related items. For example, Möller et al. (2011) measured math self-concept with three items (“*Math is one of my best subjects*”; “*In math, I do quite well*”; “*In math, I usually get good grades*”). The Cronbach’s alphas of this scale at different time points were extremely high (.90 to .91), which may indicate item redundancy (Streiner, 2003).

Descriptive statistics for these measures in the final sample are shown in Table 1. Note that mean standardised scores for math performance need not be zero as standardisation was performed within the full COOL⁵⁻¹⁸ samples, which included students who did not meet the inclusion criteria for the final sample of the present study.

Table 1

Descriptive statistics main variables

	Sex												Educational Track					
	Total		Male		Female		Low		Medium		High							
	N=843		N=394		N=449		N=329		N=235		N=279							
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD						
Grade 6:																		
Math performance ^a	0.25	0.91	0.46	0.89	0.06	0.89	-0.35	0.77	0.33	0.71	0.88	0.75						
Self-efficacy	3.71	0.58	3.80	0.56	3.63	0.59	3.50	0.57	3.78	0.57	3.90	0.51						
Grade 9:																		
Math performance ^b	0.10	0.95	0.30	0.92	-0.08	0.95	0.27	0.94	0.04	0.88	-0.06	1.00						
Self-efficacy	3.48	0.64	3.61	0.62	3.37	0.63	3.41	0.64	3.46	0.60	3.59	0.65						
Math self-concept	2.13	0.77	2.26	0.72	2.01	0.79	2.07	0.76	2.07	0.76	2.24	0.77						

Notes. ^a Standardised within full COOL⁵⁻¹⁸ Grade 6 sample (N=11,609); ^b Standardised within full COOL⁵⁻¹⁸ Grade 9 sample (N=21,384).

2.4 Analysis

Analyses were performed in IBM SPSS Statistics 20[®] ($\alpha = .05$). Preliminary GLM analyses with posthoc comparisons (Bonferroni correction) were performed to establish the extent to which between-subjects differences (i.e., *track* and *sex*) and within-subjects temporal differences (i.e., between Grade 6 and Grade 9) were present for the main variables (i.e., standardised *math performance*, *self-efficacy*,



math self-concept). *Age* was included as a covariate. Note that the temporal analysis could not be performed for *math self-concept* as it was only measured in Grade 9.

For the main analysis, a multiple mediator model² determined the extent to which the effect of *Grade 6 math performance* on *Grade 9 math performance* is mediated by *self-efficacy* and *math self-concept*. This model³ is depicted in Figure 1, assuming the direction of effects between math performance, self-efficacy and math self-concept presented in the Introduction. As multicollinearity could affect the outcomes of the analysis (Hayes, 2013) and be particularly misleading when comparing effects of self-efficacy and self-concept (Marsh, Dowson, Pietsch, & Walker, 2004), Variance Inflation Factors (VIF) and tolerances were first calculated. All VIFs were below 2.5 and all tolerances were above 0.40, indicating absence of multicollinearity. Then, Hayes' (2013) bootstrapping method⁴ was used to estimate the indirect effects of the hypothesised mediators with age as a covariate as well as confidence intervals for these effects. An indirect effect is significant if the 95% confidence interval does not contain zero. Effect size was calculated as the ratio of the indirect effect to the total effect of *Grade 6 math performance* on *Grade 9 math performance*. Simple contrasts between each pair of proposed mediators identified the most influential mediator overall.

Finally, moderated mediation analyses tested whether the strength of the indirect effects was conditional on sex and/or track. The conditional indirect effect of a specific mediator estimates the indirect effect of that mediator at specified values of the moderator. For dichotomous moderators (e.g., sex), these values represent the two groups. For moderation by track, conditional indirect effects were estimated for the (a) low versus medium tracks; (b) low versus high tracks; and (c) medium versus high tracks. The so-called Index of Moderated Mediation (IMM) tests the equality of the conditional indirect effects in the groups being compared. When the index is not significant, these effects are equivalent.

3. Results

3.1 Preliminary analyses: between-subjects and temporal differences

There was a large effect of *track*, a medium-large effect of *sex* and a small effect of *age* (Table 2). Males scored higher than females on all variables. *Tracks* also differed on all variables. In Grade 6, *self-efficacy* and *math performance* were lowest in the lowest track and *math performance* was highest in the highest track (all $p_{Bonf} < .001$). In Grade 9, *self-efficacy* and *math self-concept* were higher in the highest track than the lowest track ($p_{Bonf} = .002$ and $.02$ respectively). Although mean standardised scores (standardised relative to school and track) could be expected to be zero in all tracks, *math performance* was higher in the lowest track ($p_{Bonf} < .03$). This apparent anomaly is due to the exclusion of delayed students, most of whom were in the lowest track and had lower math performance than the other low track students. Consequently, the mean of the final low track sample was higher than zero. This has no further significance for the study findings. *Age* did not affect *self-efficacy* but did affect *math self-concept* and both *math performance* measures (in Grade 6 and Grade 9): these variables were lower for older students ($r = -.10, -.12$ and $-.11$, respectively; $p < .01$).

Temporal differences took the form of two *time* x *track* interactions: for *self-efficacy* ($F(2,836) = 11.89, p < .001, \eta_p^2 = .03$), with the lowest track showing the smallest decline, and for *math performance* ($F(2,836) = 250.87, p < .001, \eta_p^2 = .38$). In the lowest track, *math performance* in Grade 9 was higher than in Grade 6, while the reverse was true for the two higher tracks. This is consistent with the shift in reference group: many lower ability students have higher scores in Grade 9 relative to students of similar ability than in Grade 6 relative to students of all ability levels, while the converse is true for higher ability students.



Table 2
Between-subjects comparisons main variables

	Wilks' λ	F	($df1, df2$)	p	η_p^2
SEX	.88	22.21	(5,832)	< .001	.12
TRACK	.52	64.15	(10,1664)	< .001	.28
SEX*TRACK	.99	0.53	(10,1664)	.87	.00
AGE (covariate)	.98	3.48	(5,832)	.004	.02
SEX:					
Self-efficacy G6		24.22	(1,836)	< .001	.03
Self-efficacy G9		32.63	(1,836)	< .001	.04
Math self-concept G9		24.72	(1,836)	< .001	.03
Math performance G6		78.79	(1,836)	< .001	.09
Math performance G9		34.91	(1,836)	< .001	.04
TRACK:					
Self-efficacy G6		44.85	(2,836)	< .001	.10
Self-efficacy G9		6.20	(2,836)	.002	.01
Math self-concept G9		4.43	(2,836)	.012	.01
Math performance G6		225.05	(2,836)	< .001	.35
Math performance G9		9.64	(2,836)	< .001	.02

3.2 Mediation analysis

The bootstrapping estimates for the multiple mediator model are presented in Table 3 and Figure 1. Moderation estimates are presented in Table 4. The total model explained 11% of variance in *Grade 9 math performance* ($F(2,840) = 60.60, p < .001$). There were significant total and direct effects of *Grade 6 math performance* on *Grade 9 math performance* and the total indirect effect through the hypothesised mediators was also significant. *Grade 6 self-efficacy* and *Grade 9 math self-concept* each uniquely mediated the relationship between *Grade 6 math performance* and *Grade 9 math performance*, but in different directions. *Grade 9 math self-concept* was the most influential mediator, explaining 23% of the total effect, while *Grade 6 self-efficacy* had a smaller, negative relation with *Grade 9 math performance*. *Grade 9 self-efficacy* had a positive relationship to both math performance measures, but its indirect effect was not significant.

Moderation by sex. There were no sex differences in any of the indirect effects as none of the IMMs were significant, though the IMM for *Grade 6 self-efficacy* was nearly so. Specifically, the negative indirect effect of *Grade 6 self-efficacy* was significant only for females. Thus, the relation between *Grade 6 math performance* and *Grade 9 math performance* via the hypothesised mediators was similar for both sexes, but the negative relation with high self-efficacy at the end of primary school tended to affect females in particular.

Moderation by track. The indirect effect of *Grade 9 math self-concept* was significant in all tracks. The indirect effect of *Grade 6 self-efficacy* was not significant in the low or medium tracks and was borderline significant in the high track. The indirect effect of *Grade 9 self-efficacy* was significant only in the low track. The IMMs indicated one difference between tracks: the indirect effect of *Grade 9 self-efficacy* was greater in the low track than in the medium track.



Table 3
Bootstrapping results mediation analysis

	Estimate	Boot SE	ES	95% CI	B	SE	t	p
Total effect (c path)	0.33	0.03					10.48	<.001
Direct effect (c' path)	0.28	0.03					8.30	<.001
Age (covariate)					-0.23	0.11	-2.21	.03
Indirect effects:								
Total indirect	0.06	0.02	0.17	0.02 - 0.10				
Self-efficacy G6	-0.03	0.01	0.09	-0.06 - -0.00				
a path					0.22	0.02	10.37	<.001
b path					-0.14	0.06	-2.37	.02
Self-efficacy G9 ^a	0.01	0.01	0.03	-0.00 - 0.02				
a path					0.11	0.02	5.01	<.001
b path					0.08	0.05	1.50	.13
Math self-concept G9	0.08	0.01	0.23	0.05 - 0.11				
a path					0.24	0.03	8.13	<.001
b path					0.33	0.05	7.30	<.001
Contrasts:								
M1-M2	-0.04	0.02		-0.07 - -0.01				
M1-M3	-0.11	0.02		-0.15 - -0.07				
M2-M3	-0.07	0.02		-0.10 - -0.04				

Notes. ^a When math self-concept G9 is omitted, estimates for the *b* path of self-efficacy G9 ($B=0.22, SE=0.05, t=3.96, p<.001$) and the indirect effect of self-efficacy G9 ($Est=0.02, SE=0.01, ES=0.07, CI=0.01-0.04$) are significant; 5000 bootstrap samples; $\alpha = .05$; *ES* (effect size) = magnitude(indirect effect/total effect); M1 = Self-efficacy G6; M2 = Self-efficacy G9; M3 = Math self-concept G9; estimated values are rounded to 2 decimal places (e.g., -0.0016 is reported as -0.00).

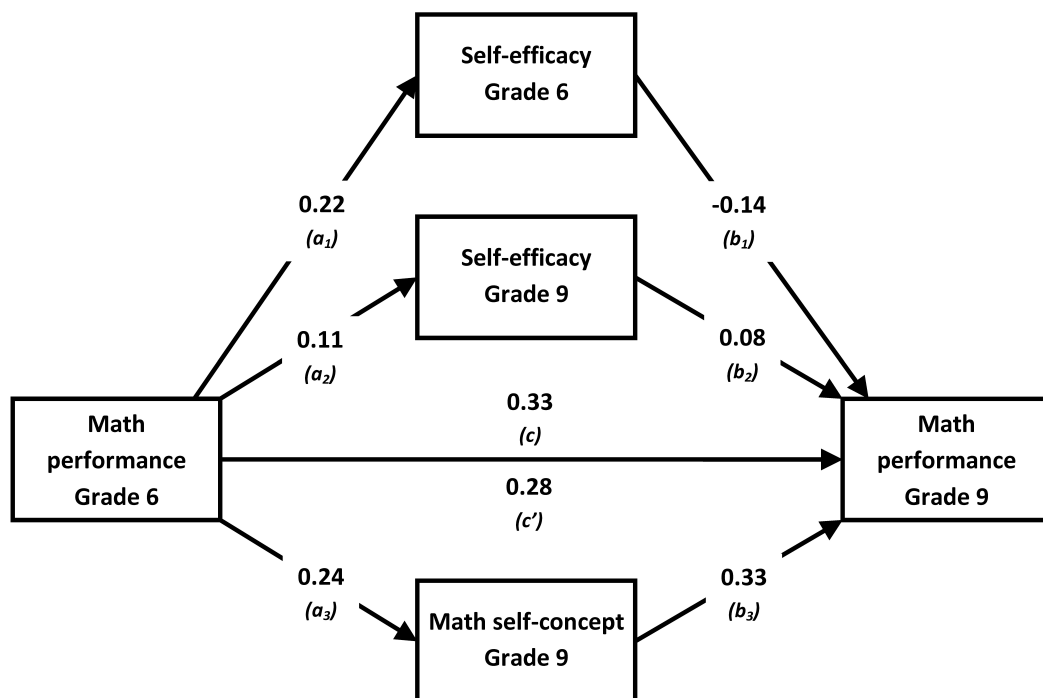


Figure 1. Multiple mediator model with bootstrapping estimates for indirect, direct and total effects.



Table 4

Bootstrapping results moderated mediation analysis conditional indirect effects

	Moderation by sex											
	Males				Females				IMM			
	Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI	
Self-efficacy G6	-0.01	0.02	-0.04	- 0.03	-0.05	0.02	-0.09	- -0.02	-0.05	0.03	-0.10	- 0.00
Self-efficacy G9	0.01	0.01	-0.00	- 0.02	0.00	0.01	-0.01	- 0.02	-0.00	0.01	-0.02	- 0.02
Math self-concept G9	0.05	0.02	0.02	- 0.10	0.08	0.02	0.05	- 0.13	0.03	0.03	-0.02	- 0.08
	Moderation by track											
	Low Track				Medium Track				High Track			
	Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI	
Self-efficacy G6	-0.01	0.02	-0.04	- 0.03	0.01	0.02	-0.02	- 0.04	-0.02	0.01	-0.05	- ±0.00
Self-efficacy G9	0.02	0.01	0.00	- 0.06	-0.01	0.01	-0.04	- 0.01	0.01	0.01	-0.01	- 0.04
Math self-concept G9	0.06	0.02	0.03	- 0.12	0.09	0.03	0.04	- 0.16	0.10	0.03	0.05	- 0.17
	IMM: Low versus Medium				IMM: Low versus High				IMM: Medium versus High			
	Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI	
	Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI		Estimate	Boot SE	95% CI	
Self-efficacy G6	-0.01	0.02	-0.06	- 0.03	0.02	0.02	-0.03	- 0.06	-0.03	0.02	-0.07	- 0.01
Self-efficacy G9	0.03	0.02	0.00	- 0.07	0.01	0.02	-0.02	- 0.05	0.01	0.01	-0.01	- 0.05
Math self-concept G9	-0.02	0.04	-0.10	- 0.05	-0.04	0.04	-0.12	- 0.03	0.02	0.04	-0.07	- 0.10

Notes. 5000 bootstrap samples; $\alpha = .05$; estimated values are rounded to 2 decimal places (e.g., -0.0049 is reported as -0.00).



4. Discussion

This study investigated the extent to which self-beliefs mediate the relation between math performance at the end of primary school (i.e., Grade 6) and the end of lower secondary school (i.e., Grade 9) in a highly differentiated early tracking educational system. The study involved 843 typically-developing students who participated in a large-scale, nationally representative, longitudinal cohort study in the Netherlands.

In interpreting the results, it is important to note that self-beliefs are shaped by comparisons with relevant reference groups (Möller et al., 2009; Möller et al., 2011; Schunk & Meece, 2006) and that math performance was standardised on the same basis. While Grade 6 students compare themselves to classmates of all ability levels (i.e., a heterogeneous reference group), the highly differentiated tracking structure of Dutch secondary education means that Grade 9 students, who are established in ability-homogeneous tracks, compare themselves to classmates in the same track as themselves. The corresponding change in reference group is likely over time to depress self-beliefs as well as relative math performance in higher tracks and increase them in lower tracks (Chmielewski et al., 2013; Liu et al., 2005; Marsh, 1991; Marsh & Hau, 2003). Indeed, exactly this pattern was found for math performance and - despite a general decline in self-efficacy from Grade 6 to Grade 9 - the lowest track showed a much smaller decline than the other two tracks.

Self-efficacy in Grade 6 and math self-concept in Grade 9 both uniquely mediated the relation between math performance in Grade 6 and in Grade 9, but self-efficacy in Grade 9 only added to the mediation effects in the lowest track. It should be noted that the mediation analysis method used here focuses on the unique contribution of each proposed mediator. Although there was no excessively high relation between the measures of self-efficacy and math self-concept in Grade 9, the existing degree of overlap clearly diminished the unique contribution of the former when the latter was taken into account (see Note ^a of Table 3).

Math self-concept was the most influential mediator, explaining nearly a quarter of the total effect of math performance in Grade 6 on math performance in Grade 9. The finding that math-specific self-beliefs (here, math self-concept) are more influential than general self-beliefs (here, self-efficacy) is consistent with previous research (Bong & Skaalvik, 2003; Valentine et al., 2004). Although causality cannot be determined from these data even with the longitudinal design, the findings suggest that higher math performance at the end of primary school may positively influence math self-concept which, in turn, may be conducive to math performance in lower secondary school. This is in line with previous research demonstrating reciprocal effects between math self-concept and performance, which shows that self-concept influences outcomes (thus, performance is improved by enhancing self-concept) and outcomes influence self-concept (thus, self-concept is enhanced by developing stronger skills) (Marsh & Martin, 2011; Möller et al., 2011).

Unexpectedly, higher self-efficacy in Grade 6 was negatively related to Grade 9 math performance in the highest track and for girls. With the same caveat regarding causality, this could mean that, when these students are confident about their academic abilities at the end of primary school, this may lead to lower math performance at the end of lower secondary school. These findings run counter to the large body of research indicating that self-efficacy has a positive influence on performance (Ferla et al., 2009; Schunk & Meece, 2006; Skaalvik & Skaalvik, 2006; Valentine et al., 2004).

Several explanations are plausible. As discussed in the Introduction, self-efficacy is shaped by several factors, including repeated successes or failures as well as appraisals by significant others. Thus, students who have completed primary school with ease - evidenced by repeated successes and reinforced by parents and teachers - may enter secondary school expecting to succeed at academic tasks. This could particularly be the case for high ability students, who are often successful in primary school with comparatively little effort. However, these students may have difficulty changing this approach in secondary school, for example spending less time on schoolwork than is necessary (cf. Vancouver & Kendall, 2006). Given the more exacting demands and conditions of secondary school - particularly in higher tracks - this approach is likely to produce lower performance.



Additionally, disparities between learning environments in primary and secondary school could mean that learning strategies that have served well and brought success in primary school may be less effective - or even counterproductive - in secondary school. Thus, students who persist in using such strategies could be at a disadvantage when dealing with schoolwork in secondary school. For example, students who habitually make use of rote-learning strategies (e.g., for learning multiplication tables) or standard algorithms for problem solving are likely to encounter difficulties when required to master concepts and solve more complex, novel problems in secondary school (Mayer, 2002). Notably, students with unrealistically high self-efficacy are often overconfident of their study methods and are unwilling to change them (Schunk & Pajares, 2004).

Furthermore, students who enter secondary school believing they will be successful face a harder 'reality check' when confronted with more demanding environments. This may produce distress that diverts attention away from learning and towards re-establishing well-being (Boekaerts, 2006). Initial problems encountered after school transition could set students on a downward path that they may not easily recover from. In any case, higher self-efficacy at the end of primary school may not necessarily be a protective factor if not appropriately managed when students move to secondary school.

Previous research reported sex differences in math-related self-beliefs (Else-Quest et al., 2010; Herbert & Stipek, 2005; Ireson & Hallam, 2009; Jacobs et al., 2002; OECD, 2013; Preckel et al., 2008; Schunk & Meece, 2006). In the present study, boys also had higher self-beliefs than girls but the patterns of relationships between self-beliefs and math performance were largely similar for both sexes. Nonetheless, the negative effect of Grade 6 self-efficacy on later math performance was significant only for girls, suggesting that the mechanisms proposed above could be less influential for boys, at least in typically-developing students. Boys have been reported to have a more positive adaptation to secondary school than girls, who are more susceptible to stress and distress during this period (Akos & Galassi, 2004; Cauley & Jovanovich, 2006). Furthermore, gender differences in mathematical problem solving strategies have been found, with girls having a greater propensity for following rules and standard algorithms (Leedy, LaLonde, & Runk, 2003; Zhu, 2007). As noted, though these strategies may bring success in primary school, they may not be conducive to more complex mathematical thinking and learning later on.

5. Future research

This study has a number of strengths that contribute to understanding the relation between self-beliefs and math performance: specifically, the large-scale longitudinal design, the use of validated self-report and performance measures, and the inclusion of students' external frames of reference (i.e., peer group comparisons). Nonetheless, certain issues not addressed here should be investigated in future research.

The negative relation between high self-efficacy at the end of primary school and later math performance was not significant for typically-developing boys. However, this relation could be stronger in underachieving or failing (i.e., delayed) boys. Boys are known to overestimate their capabilities (Pajares, 2002) and are also overrepresented among underachieving students and school dropouts (Driessen & Van Langen, 2010; Lamb, Markussen, Teese, Sandberg, & Polesel, 2011). It seems likely that unrealistic self-beliefs could contribute to these outcomes. Thus, the mediating effects of self-beliefs on performance in delayed students should be examined in future research.

Furthermore, math self-concept was not measured in Grade 6. Assuming a degree of overlap between self-efficacy and math self-concept in Grade 6, as in Grade 9, it would be of interest to isolate the effects of self-efficacy in Grade 6 when a concurrent measure of math self-concept is included.

Additional longitudinal studies with repeated measurements are needed to confirm whether the effects found here reflect causal influences. As it is often argued that enhancement of self-beliefs should be one of the key goals of education (Marsh & Martin, 2011; Möller et al., 2009; OECD, 2013; Schunk & Meece, 2006), it is important to determine their impact in educational systems with highly differentiated



early tracking. The present findings suggest that, in these systems, students' self-efficacy beliefs may need to be managed during the transition between primary school and the early years of secondary school. If initiatives to improve self-beliefs do not regard the realities that students face and their ability to adapt learning strategies to different environments, this could be detrimental to performance. In fact, unrealistically high self-beliefs have been linked to lower performance (Chiu & Klassen, 2010; Vancouver & Kendall, 2006).

Finally, while the study took account of students' external frames of reference, an internal comparison process is also recognised in the literature, whereby students compare their own achievements across several domains. These comparisons may attenuate or inflate self-concept in a particular domain, independent of actual performance (Möller et al., 2009; Möller et al., 2011; Skaalvik & Skaalvik, 2002). Future research including both frames of reference would complement other work investigating these issues in early tracking systems (e.g., Möller et al., 2009; Möller et al., 2011).

Keypoints

- Self-beliefs mediate math performance between primary and lower secondary school in a highly differentiated early tracking educational system
- Math self-concept explains a quarter of the total effect of earlier math performance on later math performance
- Self-efficacy at the end of primary school has a negative relation with later math performance, particularly for girls and high-track students
- High self-efficacy may not necessarily be a protective factor in highly differentiated early tracking educational systems

Acknowledgments

The authors thank the creators of the COOL⁵⁻¹⁸ datasets, particularly Greetje van der Werf and Hans Kuypers. The datasets were obtained from the Data Archiving Network Services (DANS) website (<http://www.dans.knaw.nl>).

COOL⁵⁻¹⁸ (2007/8):

- SCO-Kohnstamm Instituut Amsterdam; ITS Radboud Universiteit Nijmegen; CITO Arnhem; GION RU Groningen
- Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2007; Eerste meting basisonderwijs 2007 (2007-09-01, 2008-04-30)
- Persistent identifier: urn:nbn:nl:ui:13-icz-r75

COOL⁵⁻¹⁸ (2010/11):

- GION RU Groningen, CITO Arnhem, SCO-Kohnstamm Instituut Amsterdam, ITS Radboud Universiteit Nijmegen
- Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Voortgezet Onderwijs Klas 3 - 2010/11 (2012-11-09)
- Persistent identifier: urn:nbn:nl:ui:13-y9jp-e0



References

- Akos, P. & Galassi, J. P. (2004). Gender and race as variables in psychosocial adjustment to middle and high school. *The Journal of Educational Research*, *98*, 102-108. doi:10.3200/JOER.98.2.102-108
- Anderman, E. M., Urdan, T., & Roeser, R. (2003). The patterns of adaptive learning survey: History, development and psychometric properties. *Indicators of Positive Development Conference*. March 12-13, 2003, Washington, D.C.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Boekaerts, M. (2006). Self-regulation and effort investment. In K. A. Renninger & I. E. Sigel (Eds.), *Handbook of child psychology : Vol. 4. Child psychology in practice* (6th ed., pp. 345-377). Hoboken, NJ: John Wiley & Sons.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*, 1-40. doi: 10.1023/A:1021302408382
- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *17*, 141-159. doi:10.1080/13803611.2011.597112
- Brophy, J. (2006). *Grade repetition. Education policy series No. 6*. Belgium/Paris: International Academy of Education and International Institute for Educational Planning, UNESCO.
- Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology*, *81*, 78-96. doi:10.1348/2044-8279.002004
- Cauley, K. M., & Jovanovich, D. (2006). Developing an effective transition program for students entering middle school or high school. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *80*, 15-25. doi:10.3200/TCHS.80.1.15-25
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*, 2-17. doi:10.1016/j.learninstruc.2008.11.002
- Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type: An international comparison of students' mathematics self-concept. *American Educational Research Journal*, *50*, 925-957. doi:10.3102/0002831213489843
- Crone, E. A., Wendelken, C., Van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*, 55-66. doi:10.1111/j.1467-7687.2008.00743.x
- Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & Van der Veen, I. (2009). *Cohortonderzoek COOL⁵⁻¹⁸: Technisch rapport basisonderwijs, eerste meting 2007/08* [Cohort Study COOL⁵⁻¹⁸: Technical report primary education, first cohort measurement 2007/08]. Nijmegen: ITS/Amsterdam: SCO-Kohnstamm Instituut, The Netherlands.
- Driessen, G., & Van Langen, A. (2010). *De onderwijsachterstand van jongens. Omvang, oorzaken en interventies* [The educational disadvantage of boys. Extent, causes and interventions]. Nijmegen, The Netherlands: ITS - Radboud University Nijmegen.
- Dumontheil, I., Houlton, R., Christoff, K. & Blakemore, S-J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, *13*, F15-F24. doi:10.1111/j.1467-7687.2010.01014.x
- Else-Quest, N-M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103-127. doi:10.1037/a0018053
- Fenzel, L. M. (2000). Prospective study of changes in global self-worth and strain during transition to middle school. *Journal of Early Adolescence*, *20*, 93-116. doi:10.1177/0272431600020001005
- Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, *19*, 499-505. doi:10.1016/j.lindif.2009.05.004
- Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*,



898-915. doi:10.1177/0013164498058006003

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hanushek, E. A., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, *116*(510), C63-C76. doi:10.1111/j.1468-0297.2006.01076.x
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.
- Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, *26*, 276-295. doi:10.1016/j.appdev.2005.02.007
- Hoogeveen, L., Van Hell, J. G., & Verhoeven, L. (2009). Self-concept and social status of accelerated and nonaccelerated students in the first 2 years of secondary school in the Netherlands. *Gifted Child Quarterly*, *53*, 50-67. doi:10.1177/0016986208326556
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*, 505-528. doi:10.1016/j.jsp.2011.07.001
- Ireson, J., & Hallam, S. (2009). Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction*, *19*, 201-213. doi:10.1016/j.learninstruc.2008.04.001
- Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, *1*(3), 33-58. doi:10.1257/app.1.3.33
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509-527. doi:10.1111/1467-8624.00421
- Lamb, S., Markussen, E., Teese, R., Sandberg, N., & Polesel, J. (2011). *School dropout and completion: International comparative studies in theory and policy*. Dordrecht, The Netherlands: Springer Science+Business Media B.V.
- Leedy, M. G., LaLonde, D., & Runk, K. (2003). Gender equity in mathematics: Beliefs of students, parents, and teachers. *School Science and Mathematics*, *103*, 285-292. doi:10.1111/j.1949-8594.2003.tb18151.x
- Liu, W. C., Wang, C. K., & Parkins, E. J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology*, *75*, 567-586. doi:10.1348/000709905X42239
- Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal Summer*, *28*, 445-480. doi:10.3102/00028312028002445
- Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, *96*, 518-522. doi:10.1037/0022-0663.96.3.518
- Marsh, H. W., & Hau, K-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*, 364-376. doi:10.1037/0003-066X.58.5.364
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*, 59-77. doi:10.1348/000709910X503501
- Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal*, *37*, 739-763. doi:10.1080/01411926.2010.490874
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, *41*, 226-232. doi:10.1207/s15430421tip4104_4
- Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., et al., (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.



- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*, 1129-1167. doi:10.3102/0034654309337522
- Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal, 48*, 1315-1346. doi:10.3102/0002831211419649
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.
- OECD. (2012). *Equity and quality in education: Supporting disadvantaged students and schools*. OECD Publishing. doi:10.1787/9789264130852-en
- OECD. (2013). *PISA 2012 results: Ready to learn - Students' engagement, drive and self-beliefs (Volume III)*. PISA, OECD Publishing. doi:10.1787/9789264201170-en
- Osterman, K. F. (2000). Students' need for belonging in the school community. *Review of Educational Research, 70*, 323-367. doi:10.3102/00346543070003323
- Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice, 41*, 116-125. doi:10.1207/s15430421tip4102_8
- Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. J. Riding & S. G. Rayner (Eds.), *Self perception* (pp. 239-265). Westport, CT: Ablex Publishing.
- Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students: Comparing girls' and boys' achievement, self-concept, interest and motivation in mathematics. *Gifted Child Quarterly, 52*, 146-159. doi:10.1177/0016986208315834
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*, 151-161. doi:10.1177/0146167201272002
- Sakiz, G., Pape, S. J., & Hoy, A. W. (2012). Does perceived teacher affective support matter for middle school students in mathematics classrooms? *Journal of School Psychology, 50*, 235-255. doi:10.1016/j.jsp.2011.10.005
- Schunk, D. H., & Meece, J. L. (2006). Self-efficacy development in adolescence. In F. H. Pajares & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 71-96). Greenwich, CT: Information Age Publishing.
- Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (pp. 115-138). Greenwich, CT: Information Age Publishing.
- Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist, 37*, 233-244. doi:10.1207/S15326985EP3704_3
- Skaalvik, E. M., & Skaalvik, S. (2006). Self-concept and self-efficacy in mathematics: Relation with mathematics motivation and achievement. In A. P. Prescott (Ed.), *The concept of self in education, family and sports* (pp. 51-74). New York: Nova Science Publishers.
- Somerville, L. H., & Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology, 20*, 236-241. doi:10.1016/j.conb.2010.01.006
- Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*, 80-90. doi:10.1016/j.lindif.2008.05.004
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99-103. doi:10.1207/S15327752JPA8001_18
- Urdan, T., & Midgley, C. (2003). Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence. *Contemporary Educational Psychology, 28*, 524-551. doi:10.1016/S0361-476X(02)00060-7
- Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology, 44*, 331-349. doi:10.1016/j.jsp.2006.04.003



- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*, 111-133. doi:10.1207/s15326985ep3902_3
- Vancouver, J. B., & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology*, *91*, 1146-1153. doi:10.1037/0021-9010.91.5.1146
- Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: Foundations, recent developments, and applications* (pp. 215-238). New York: Springer-Verlag.
- Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, *8*, 187-203.
- Zijsling, D., Keuning, J., Naayer, H., & Kuyper, H. (2012). *Cohortonderzoek COOL⁵⁻¹⁸: Technisch rapport meting VO-3 in 2011* [Cohort Study COOL⁵⁻¹⁸: Technical report Grade 9 cohort measurement in 2011]. Groningen, The Netherlands: GION.

Appendix: Methodological footnotes

¹ The COOL⁵⁻¹⁸ study was commissioned by the Netherlands Organisation for Scientific Research and the Ministry of Education, Culture and Science, and was carried out by a broad consortium of research and assessment organisations in the Netherlands. Full descriptions of participants, methods and procedures are provided in the technical reports (Driessen, Mulder, Ledoux, Roeleveld, & Van der Veen, 2009; Zijsling, Keuning, Naayer, & Kuyper, 2012).

² A mediation model is a type of Structural Equation Model, referring to a sequence of relations in which an independent variable affects a dependent variable by influencing intervening (i.e., mediator) variables. The order of the variables must be established on theoretical, logical or procedural grounds (Hayes, 2013).

³ The a_i paths represent the effect of *Grade 6 math performance* on the proposed mediators. The b_i paths represent the effect of the proposed mediators on *Grade 9 math performance*, partialling out the effect of *Grade 6 math performance*. Path c represents the total effect of *Grade 6 math performance* on *Grade 9 math performance* and path c' represents the direct effect of *Grade 6 math performance* on *Grade 9 math performance* after controlling for the proposed mediators. The specific indirect effect of *Grade 6 math performance* on *Grade 9 math performance* through a particular mediator (i.e., the unique ability of the mediator to mediate the effect of *Grade 6 math performance* on *Grade 9 math performance* conditional on the other mediators) is the product of the two paths linking *Grade 6 math performance* to *Grade 9 math performance* via that mediator (i.e., $a_i * b_i$). The total indirect effect of *Grade 6 math performance* on *Grade 9 math performance* is the sum of the specific indirect effects. The total effect of *Grade 6 math performance* on *Grade 9 math performance* (path c) is the sum of the direct effect and all of the specific indirect effects.

⁴ The bootstrapping method is implemented in Hayes' PROCESS macro (obtained from <http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html>). A strength of this procedure is that it does not make assumptions about the sampling distribution of the indirect effects or force choices about estimation or constraint of residual covariances. It resamples thousands of times from the dataset and estimates the indirect effects in each resample, thereby providing an empirical approximation of and confidence intervals for these effects. Bias-corrected confidence intervals were used, as indirect effects usually have a skewed distribution. A heteroscedasticity-consistent standard error estimator was used, which reduces the likelihood that inference validity is compromised by any potential violation of homoscedasticity. Model 4 in the PROCESS macro was used to estimate the indirect effects of the hypothesised mediators. Model 59 was used for the moderated mediation analyses.