Mathematical Thinking, Learning and Performance:

Reed, H.C.

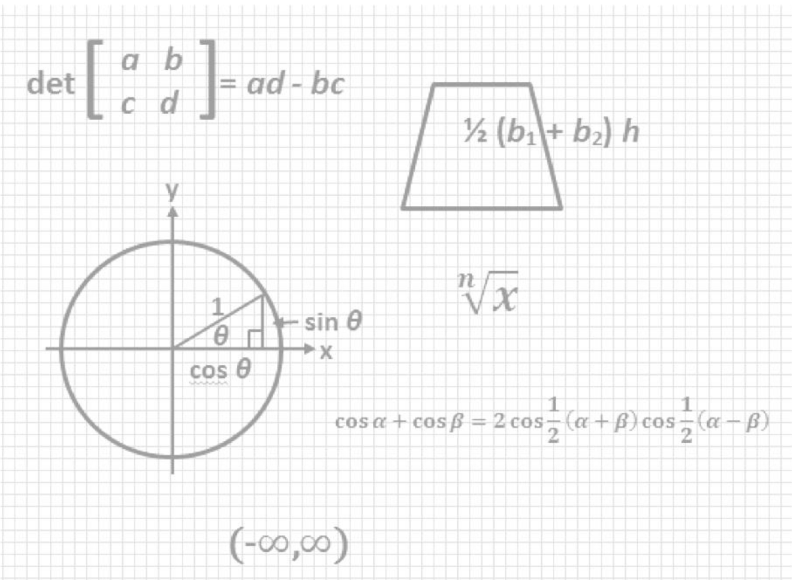
2014

7

Abstract and relational thinking and mathematical performance at the end of lower secondary school

Helen C. Reed, Paul A. Kirschner and Jelle Jolles

Submitted



ABSTRACT

Mathematical thinking requires the ability to think about abstract concepts and reason about relationships between them. This large-scale study involving 3,663 students in the Netherlands investigates how these abilities relate to mathematical performance at the end of lower secondary school (Grade 9) and whether sex, socio-economic background, educational track and delay in educational progression affect this relation. Abstract and relational thinking were measured with neuropsychological tests of concept understanding and formation. Mathematical performance was measured with a validated national test. Abstract and relational thinking jointly explain 39% variance and 18% unique variance in mathematical performance; unique variance drops to 5% when educational track is accounted for. Students in higher tracks have higher mathematics performance and better abstract and relational thinking skills. Sex and delay have additional small effects, but socio-economic background is overshadowed by tracking effects. Compared to age-typical controls, age-delayed students have lower mathematical performance but equivalent abstract and relational thinking skills.

Keywords: abstract thinking; relational thinking; mathematical performance; secondary education; individual differences

INTRODUCTION

Current European debate on school success or failure (e.g., dropout) focuses on the determinants of educational progression in lower secondary school (European Commission, 2011; Eurydice, 2011). Specific individual and demographic factors - such as sex, socio-economic background, educational tracking and delay in educational progression - are thought to impact various processes that underlie student outcomes. To better understand the relation between these factors and student outcomes, it is important to consider the period of lower secondary education from the perspective of the development of specific thinking skills underlying school success. Two key skills are abstract and relational thinking, which are fundamental for many academic areas including languages, problem-solving, science and mathematics (Devlin, 2000; Dunbar & Fugelsang, 2005; Epstein, 2013; Gentner & Colhoun, 2010; Kurtz & Loewenstein, 2007; Richland, Stigler, & Holyoak, 2012).

This study investigates the relation between abstract and relational thinking skills and mathematics performance. Mathematics depends crucially on understanding and reasoning with abstract and relational concepts, as will be reviewed presently. Given worldwide concern to improve students' mathematical performance, fuelled by international assessments such as the OECD Programme for International Student Assessment (PISA; OECD, 2013a), it is surprising that little attention is paid by curriculum developers and teachers to the development of these skills.

From a developmental perspective, it is intriguing that very little progress is made in mathematical competence during the initial years of secondary school (Ryan & Williams, 2007). A possible explanation could be that secondary school mathematics requires students to think in ways that initially exceed their capabilities. Neuroimaging studies show that brain activity during abstract and relational thinking tasks involves brain networks that undergo protracted maturation throughout early to mid-adolescence (e.g., Bazargani, Hillebrandt, Christoff, & Dumontheil, 2013; Crone et al., 2009; Dumontheil, Houlton, Christoff, & Blakemore, 2010; Eslinger et al., 2009; Grossman et al., 2002; Reber, Wong, & Buxton, 2002). It is therefore conceivable that a mismatch between educational demands and the rate of maturational change in mediating brain areas could constrain performance in tasks requiring abstract and relational thinking. This could be exacerbated by the lack of attention paid to the acquisition of these skills at school. Together, this could impact mathematical performance (as well as performance in other areas) and contribute to students' success or failure in school. It is therefore important to understand the relation between abstract and relational thinking and mathematical performance in this period.

Moreover, it is likely that sex, socio-economic background, educational track and delay in educational progression influence this relation, as these factors play an important role in school outcomes in general and in mathematics performance specifically (Brophy, 2006; Buchmann, DiPrete, & McDaniel, 2008; Driessen & Van Langen, 2010; European Commission, 2011; Gamoran, 2004; Hanushek & Wößmann, 2006; Jacob & Lefgren, 2009; Jimerson, Anderson, & Whipple, 2002; Lamb, Markussen, Teese, Sandberg, & Polesel, 2011; Marks, Cresswell, & Ainley, 2006; OECD, 2011, 2012b; Schütz, Ursprung, & Wößmann, 2008; Schwerdt & West, 2012; Van de Werfhorst & Mijs, 2010). PISA assessments (OECD, 2013a, 2013b, 2013d) also confirm the influence of these factors in mathematics performance. Across OECD countries, boys outperform girls and the gap between them has widened over the past decade in several countries. Performance of more socio-economically disadvantaged students lags considerably behind that of more-advantaged students. Performance is also lower in educational systems that stratify students according to general scholastic ability (i.e., tracking), and in systems with higher rates of grade retention leading to delay in educational progression.

Abstract and relational thinking in mathematics

Mathematics has been described as the science of patterns that for the most part are represented by abstract symbolic systems (Devlin, 2000; Kieran, 2011; NCTM, 2000; Resnik, 1997). The ability to reason with the relationships within such systems allows understanding of the properties, structures and behaviours of mathematical objects and entities. For example, the abstract notation of algebra allows mathematical relationships to be described in a general way (as in: $a+b = b+a$ or $ax(bxc) = (axb)xc$) that is relevant to a wide range of mathematical topics taught at school. Consequently, the ability to handle abstraction and reason about relationships between objects or ideas is key to mathematical thinking (Devlin, 2000; Epstein, 2013; Richland et al., 2012).

Abstraction occurs when properties of a given object or collection of objects are extracted from the original(s), in order to apply them to a broader domain or to develop conceptual knowledge about certain phenomena (Harel & Tall, 1991). During this process, mental objects that are abstract representations of reality are constructed. For example, natural numbers are reifications of abstractions from counting and adding collections of objects, and Euclidean plane geometry abstracts small dots into locations without dimensions and physical lines into extensions without breadth (Epstein, 2013).

A key element of abstract thinking is the ability to understand and develop concepts derived from *categorisation* of objects, events or ideas according to certain shared principles or commonalities (Keil, 1989). Thus, in geometry, for example, shapes are first only recognised in terms of their physical appearance. Then, properties of figures begin to be discerned and are used to conceptualise classes of shapes with certain general properties. Next, interrelationships of properties both within and among figures are understood and classes or categories of figures are recognised (Van Hiele, 1986).

Relational thinking is the ability to identify connections between ideas or objects and to use these to learn more about the entities involved or to solve a problem. This ability is pivotal to mathematical understanding (Carpenter, Levi, Franke, & Zeringue, 2005; Knuth, Stephens, McNeil, & Alibali, 2006; Molina, Castro, & Ambrose, 2005; Molina & Mason, 2009; Richland et al., 2012). For example, students who do not understand that the equal sign denotes relational equivalence will typically write 11 on the line in the equation 6 + 5 = _ + 8. Students who use relational thinking, on the other hand, understand that the two expressions are related by the equal sign and that both represent the same quantity (Carpenter et al., 2005; Molina & Mason, 2009).

At the core of relational thinking is *analogical reasoning*: the ability to identify relational correspondences and similarities between a familiar domain or context and a novel one, and to use this information to solve a novel problem (Gentner & Colhoun, 2010; Richland et al., 2012). Analogical reasoning enables transfer of knowledge through the induction of generally applicable schemata covering multiple cases exhibiting the same relational properties (Bartha, 2013; Holyoak, 2012; Richland et al., 2012).

Abstract and relational thinking are inextricably connected. Abstraction can be conceived as a process of focusing on relationships between entities, regardless of their individual qualities, thereby constructing a mental object that can be expressed as an abstract model (Kotovsky & Gentner, 1996; Van Hiele, 1986). For example, general schemata expressing the relations between arithmetical operations (e.g., multiplication and division) are formed by abstracting common properties from multiple exemplars that demonstrate the (in this case, inverse) relationship between those operations (Molina et al., 2005).

The present study

The present large-scale, nationally representative study investigates the extent to which abstract and relational thinking predicts mathematical performance at the end of lower secondary school (Grade 9) in the Netherlands. Furthermore, the effects of sex, socio-economic background, educational track and delay in educational progression are determined. The Netherlands is a particularly appropriate country to investigate these issues. PISA assessments have consistently shown that Dutch 15-year old girls have lower mathematics scores than boys; moreover, girls are overrepresented at the lowest skills level and boys at the highest (OECD, 2013a). The Netherlands has above average socio-economic diversity and socio-economic status shows a clear relation to performance (OECD, 2013b). There is a high incidence of delay in educational progression: nearly 28% of 15-year old Dutch students report having repeated at least one year at school (OECD, 2013d). Furthermore, Dutch secondary education is characterised by educational tracking from early secondary school. The two highest tracks (i.e., a 6-year pre-university track and a 5-year higher general secondary education track) prepare more able students for higher education (i.e., university or college) and professional employment. Around 40% of secondary school students are served by these tracks. The remaining students receive practical training or follow a pre-vocational track, which provides practice-oriented or theory-oriented preparation for upper secondary vocational education.

We investigated the research questions in a sample of 3,663 students with Dutch nationality. The sample comprised two groups: students with age-typical educational progression (14.4-15.6 years; $N = 2{,}862$) and students with age-related delay (15.6-16.6 years; $N = 801$). The large, representative age-typical sample allows strong inferences to be drawn about effects in the population of typically developing students. A further strength of the study is that age-delayed students were analysed with a case-control design whereby they were randomly matched with age-typical controls on several background variables. This approach is indicated in the present case because of the relatively small proportion of delayed progression in the full sample (22%).

Abstract and relational thinking were measured using validated, norm-referenced tests that conform to principles for neuropsychological assessment of concept formation and reasoning described by Lezak, Howieson, Bigler and Tranel (2012).The tests involved two forms of abstract thinking (i.e., *understanding abstract concepts*; *categorisation of collections of items* based on similarities or differences) and one form of relational thinking (i.e., *identifying analogical structures*). Mathematical performance was measured with a validated national test.

In short, by examining abstract and relational thinking skills and their relation to mathematical performance at the end of lower secondary school, and by examining the effects of gender, socio-economic background, educational track and delay in educational progression, the present study has two goals. First, it elucidates general and specific patterns that could support attention to abstract and relational thinking skills in secondary school as a means to improve students' mathematics performance. Second, it provides insights into the influence of specific individual and demographic factors that are relevant for both educators and educational policymakers in current debate on the determinants of educational progression in lower secondary school.

METHODS

Data were obtained from the first cohort measurement of the COOL$^{5\text{-}18}$ study (Cohort Onderzoek Onderwijs Loopbanen; Cohort Research on Educational Careers), a large-scale, longitudinal, nationally representative study into the determinants of the cognitive and social-emotional development of children and adolescents in the Netherlands. The study was commissioned by the Netherlands Organisation for Scientific Research (NWO) and the Ministry of Education, Culture and Science and was carried out by a broad consortium of research and assessment organisations in the Netherlands. The datasets are available for secondary analysis by third-parties (e.g., the present study) and can be obtained from the website of Data Archiving Network Services (DANS); http://www.dans.knaw.nl). The technical report of the first cohort measurement, including a full description of participants, methods and procedures, can also be obtained from DANS, as well as from the study website (http://www.cool5-18.nl/voortgezetonderwijs; Zijsling, Keuning, Kuyper, Van Batenburg, & Hemker, 2009).

The first cohort measurement of the COOL$^{5-18}$ study included 8,890 ninth grade students from 81 schools across the Netherlands. Participants took several tests, including a mathematics test and tests for understanding abstract concepts, categorisation and analogical thinking (see Measurement section). Participants also completed a self-report questionnaire (e.g. school functioning, attitudes, wellbeing), parents/caregivers completed a demographic questionnaire (e.g., nationality, parental level of education (PLE)), and schools provided administrative data (e.g., age, sex, educational track). As far as can be ascertained from the technical report, missing data was not biased. The following paragraphs describe the participants, instruments and data relevant to the present study.

Participants

Participants were selected from the COOL$^{5-18}$ dataset when they had Dutch nationality and when complete data were available for sex, PLE, educational track, the mathematics test and the tests for understanding abstract concepts, categorisation and analogical thinking. In addition, students had to be aged between 14.4 and 16.6 years. This is based on the fact that Dutch secondary education uses an age-related class system based on a specific cut-off birth date. This results in a within-class age-difference of a year between the oldest and the youngest students with age-typical progression. As data collection took place from March to May, age-typical ninth grade students would be aged between 14.4 and 15.6 years. Students who had repeated a grade once would typically be a year older and therefore be aged between 15.6 and 16.6 years. Older students who may have repeated a grade more than once were excluded, as these probably differ from their classmates in several respects that could confound the results. The selected students came from 74 secondary schools.

Educational tracks were designated ‘low’ (i.e., practical training and pre-vocational education), ‘medium’ (i.e., higher general secondary education) and ‘high’ (i.e., pre-university education). The highest level of education attained by either parent/caregiver (PLE) - as an indicator of socio-economic background - was trichotomised as 'low' (highest level attained: lower vocational education), 'medium' (highest level attained: secondary vocational education) or 'high' (highest level attained: pre-university education, higher professional education or university). Participants with low PLE were subsequently excluded from analysis because they were severely underrepresented (only 7.5%).

The resulting sample comprised two groups of students: $N = 2{,}862$ age-typical students ($M_{age} = 15.1$, $SD_{age} = 0.3$) and $N = 801$ age-delayed students ($M_{age} = 16.0$, $SD_{age} = 0.3$). In the age-typical group, 47% ($N = 1{,}345$) was male and 53% ($N = 1{,}517$) female; 43% ($N = 1{,}219$) had medium PLE and 57% ($N = 1{,}643$) high PLE; 36% ($N = 1{,}035$) was in the low track, 30% ($N = 866$) was in the medium track and 34% ($N = 961$) was in the high track. In the age-delayed group, 57% ($N = 455$) was male and 43% ($N = 346$) female; 51% ($N = 405$) had medium PLE and 49% ($N = 396$) high PLE; 58% ($N = 467$) was in the low track, 27% ($N = 213$) was in the medium track and 15% ($N = 121$) was in the high track.

Instruments and data

Mathematical performance (variable: MATH, max score 100)

Participants were administered a multiple-choice mathematics test constructed by the Dutch National Institute for Educational Measurement (CITO). Test items were drawn from an item-bank of 50 items and were administered in three test versions comprising 29 or 30 multiple-choice items on arithmetic, algebra, geometry and statistics. Each item had four response options and participants had to underline their choice on an answer sheet. An example item is:

$d^2 - 6d + 5 =$

A (d - 1)(d - 5)

B (d + 1)(d + 5)

C (d - 1)(d + 6)

D (d + 1)(d - 6)

Items were analysed using the One-Parameter Logistic Model (OPLM; Verhelst & Glas, 1995) from Item Response Theory. The OPLM was used to translate raw test scores to skill-scores that were in turn 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. The bank-scores were used as the MATH measure.

Understanding abstract concepts (variable: UNDABS, max score 100)

Participants answered 30 multiple-choice items constructed by the Dutch National Institute for Educational Measurement (CITO) for which they had to choose the meaning or opposite meaning of an indicated word (abstract noun, adjective or verb) presented in sentence form. Test items were drawn from an item-bank of 60 items and were administered in three test versions. Each item had four response options. An example items is:

If someone is privileged, then…

A he [*sic*] has a better position than others.
B he has prejudices against others.
C others see him as an authority.
D others model themselves on him.

As described previously, raw scores were translated to skill-scores and bank-scores on a scale of 0 to 100% using the OPLM. The bank-scores were used as the UNDABS measure.

Categorisation (variable: CATGRS, max score 15)

Participants were administered the *categorisation* task of the Grade 9 version of the Niet Schoolse Cognitieve Capaciteiten Test (NSCCT; Non-scholastic Cognitive Capacity Test; Van Batenburg & Van der Werf, 2004). This task measures the ability to group together discriminable properties by means of a general principle or rule. The test comprised 15 multiple-choice items. For each item, three words were presented (e.g., clay, sand, mud), all belonging to a particular category. Participants had to choose from four other words (e.g., water, soil, mountain, field) the one belonging to that same category. Each correct answer was given 1 point and the total number of points was the CATGRS measure. If less than half the test items were answered, the CATGRS measure was not calculated. The internal reliability of the test was satisfactory (Cronbach's $\alpha = .77$).

Identifying analogical structures (variable: ANALOG, max score 10)

Participants were administered the *analogies* task of the Grade 9 version of the NSCCT (10 multiple-choice items). For each item, two words were presented that have a certain relationship to each other, for example 'hunger - eat'. A third word was then presented, e.g., 'thirst'. Participants had to recognise the relationship between the first two words and use it to choose from four other words (e.g., water, drink, tap, desert) the one with the same relationship to the third word. Each correct answer was given 1 point and the total number of points was the ANALOG measure. If less than half the test items were answered, the ANALOG measure was not calculated. The internal reliability of the test (Cronbach's $\alpha = .60$) was just satisfactory for large-sample, group-design psychological research (Aiken, 2000; Gregory, 2000).

Descriptive statistics for the main variables are reported in Table 7.1 and correlations are reported in Table 7.2.

Table 7.1
Descriptive statistics main variables

		MATH		UNDABS		CATGRS		ANALOG	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age-typical									
Sex:									
Male	1345	72.48	*16.26*	78.73	*14.77*	10.75	*2.79*	6.58	*1.84*
Female	1517	69.47	*15.91*	78.58	*13.11*	11.04	*2.50*	6.70	*1.74*
PLE:									
Medium	1219	66.87	*16.58*	75.23	*14.68*	10.45	*2.78*	6.31	*1.81*
High	1643	73.87	*15.13*	81.19	*12.74*	11.24	*2.49*	6.89	*1.72*
Track:									
Low	1035	58.56	*15.26*	68.74	*15.47*	9.76	*2.83*	5.66	*1.75*
Medium	866	72.65	*11.76*	81.07	*9.70*	10.95	*2.47*	6.84	*1.61*
High	961	82.56	*9.88*	87.16	*7.11*	12.09	*1.98*	7.53	*1.42*
Age-delayed									
Sex:									
Male	455	64.91	*16.85*	74.37	*16.23*	9.78	*3.05*	5.94	*2.04*
Female	346	60.14	*18.54*	73.31	*16.48*	10.35	*2.94*	6.15	*1.95*
PLE:									
Medium	405	59.02	*17.62*	70.01	*17.55*	9.62	*2.94*	5.77	*2.02*
High	396	66.77	*17.03*	77.90	*13.94*	10.44	*3.03*	6.30	*1.96*
Track:									
Low	467	54.14	*15.97*	66.30	*16.22*	9.20	*2.95*	5.33	*1.90*
Medium	213	72.24	*12.09*	82.50	*8.93*	10.67	*2.92*	6.68	*1.67*
High	121	79.96	*10.28*	88.18	*7.49*	12.11	*1.92*	7.63	*1.58*

Table 7.2
Bivariate correlations main variables

	N	MATH-UNDABS	MATH-CATGRS	MATH-ANALOG	UNDABS-CATGRS	UNDABS-ANALOG	CATGRS-ANALOG
				Age-typical			
Sex:							
Male	1345	.55***	.39***	.45***	.37***	.40***	.41***
Female	1517	.54***	.41***	.43***	.38***	.43***	.34***
PLE:							
Medium	1219	.52***	.38***	.44***	.36***	.40***	.35***
High	1643	.52***	.37***	.40***	.35***	.38***	.37***
Track:							
Low	1035	.36***	.23***	.24***	.27***	.23***	.28***
Medium	866	.19***	.19***	.22***	.18***	.21***	.24***
High	961	.24***	.23***	.23***	.16***	.24***	.25***
				Age-delayed			
Sex:							
Male	455	.48***	.39***	.42***	.35***	.42***	.40***
Female	346	.65***	.44***	.50***	.49***	.51***	.39***
PLE:							
Medium	405	.52***	.39***	.42***	.43***	.41***	.39***
High	396	.57***	.37***	.44***	.34***	.48***	.39***
Track:							
Low	467	.41***	.29***	.29***	.31***	.31***	.29***
Medium	213	.03	.12	.09	.14*	.15*	.21**
High	121	.26**	.21*	.28**	.29**	.32***	.47***

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Analysis

Analyses were performed in IBM SPSS Statistics 20® with an α of .01 because of the large sample-size. Based on Mahalanobis distances to detect multivariate outliers, 34 participants with unusual combinations of scores on the four main variables were excluded from analysis. Scatterplots of the standardised residuals as a function of standardised predicted values were examined to confirm normality, linearity, and homoscedasticity of residuals. A further eight outliers were identified having standardised residuals of abnormally large magnitude. The final dataset for analysis contained $N = 2{,}836$ age-typical students and $N = 785$ age-delayed students.

Analyses were performed separately for age-typical students and age-delayed students. For the age-typical students, group differences in the four main constructs (i.e., MATH, UNDABS, CATGRS and ANALOG) for the background variables (i.e., sex, PLE, educational track) were tested with an omnibus 2x2x3 MANOVA followed by univariate ANOVAs for the significant effects and posthoc comparisons (Bonferroni correction). Then, multiple regression analysis investigated the extent to which understanding abstract concepts (UNDABS), categorisation (CATGRS) and identifying analogical structures (ANALOG) predict mathematical performance (MATH), and whether sex, PLE and educational track impact this relationship. The predictors were added sequentially to the regression model (Model 1: UNDABS, CATGRS, ANALOG; Model 2: sex (reference group: males); Model 3: PLE (reference group: medium PLE); Model 4: educational track (reference group: low track)).

Age-delayed students were analysed with a case-control design. They were randomly matched with age-typical students on sex, PLE, educational track and school. If no match was found on this combination of variables, geographical region was used instead of school. Ultimately, 781 (99%) of the age-delayed students could be matched with age-typical controls. The analyses therefore used a sample comprising N=1,562 students, equally split between age-typical and age-delayed students. Group differences in the four main constructs were tested with a 2x2x2x3 (i.e., educational progression (age-typical or age-delayed), sex, PLE, educational track) MANOVA and the regression analysis included delay as an additional predictor (reference group: age-typical).

RESULTS

Age-typical students

There were between-groups differences on the main variables for *sex* (Wilks' $\lambda = .97$, $F(4{,}2821) = 22.18$, $p < .001$, $\eta_p^2 = .03$) and *track* (Wilks' $\lambda = .55$, $F(8{,}5642) = 242.21$, $p < .001$, $\eta_p^2 = .26$) but not for *PLE*. Males scored higher than females on MATH ($F(1{,}2824) = 74.04$, $p < .001$, $\eta_p^2 = .03$) and UNDABS ($F(1{,}2824) = 13.01$, $p < .001$, $\eta_p^2 = .00$). There were no sex differences on CATGRS or ANALOG. Scores on all four variables were highest in the high track and lowest in the low track (MATH: $F(2{,}2824) = 798.18$, $p < .001$, $\eta_p^2 = .36$; UNDABS: $F(2{,}2824) = 565.05$, $p < .001$, $\eta_p^2 = .29$; CATGRS: $F(2{,}2824) = 178.14$, $p < .001$, $\eta_p^2 = .11$; ANALOG: $F(2{,}2824) = 287.67$, $p < .001$, $\eta_p^2 = .17$; all $p_{Bonf} < .001$). There were no 2- or 3-way interactions (i.e., *sex* x *track*, *PLE* x *track*, *sex* x *PLE* or *sex* x *PLE* x *track*).

The results of the regression analyses are presented in Table 7.3. For all models, UNDABS, CATGRS and ANALOG each significantly and positively predicted MATH, with UNDABS being the most important predictor of the three. When track was not taken into account, a 1 point increase in each of these three variables (on scales of 1-100, 1-15 and 1-10 respectively) improved MATH by approximately ½, 1 and 2 points. Together, they accounted for 18% unique variance and explained 39% of the total variance in mathematical performance. *Sex* negatively predicted MATH over and above UNDABS, CATGRS and ANALOG and accounted for an additional 1% of unique variance: females scored 3 points lower than males. Students with high *PLE* scored over 2 points higher than the average reference group student. This did not account for additional unique variance, however.

When *track* was included, this had an important effect. Relative to the average reference group student, students in the medium track scored 8¾ points higher on MATH, while students in the high track scored 15½ points higher. Total explained variance rose to 50%, of which 14% was uniquely accounted for by track. At the same time, the unique variance explained by UNDABS, CATGRS and ANALOG decreased to 5% and the gender gap increased further. The effect of having a high PLE became non-significant; thus, PLE is not meaningfully differentiable from the effects of track.

To investigate these last two points, post-hoc tests (Chi-square test and Fisher's exact test) were performed to compare the proportions of sex and PLE across tracks (sex: $\chi^2(2) = 12.24$, $p = .002$; PLE: $\chi^2(2) = 303.83$, $p < .001$). There was a higher proportion of males in the low track and a higher proportion of females in the higher track. The additional decline in females' MATH scores when the low track was the reference group thus suggests that the gender gap is smaller in the low track than in the other tracks (this was confirmed with a *sex* x *track* profile plot, though the interaction was not significant). There was a higher proportion of participants with medium PLE in the low track and proportionally more high-PLE participants in the medium and high tracks. This could account for the overshadowing of PLE differences by track.

Table 7.3

Multiple regression for mathematical performance: age-typical students

	Model 1				Model 2				Model 3				Model 4			
	B	*SE*	*β*	sr^2	*B*	*SE*	*β*	sr^2	*B*	*SE*	*β*	sr^2	*B*	*SE*	*β*	sr^2
Intercept	9.51	1.50			11.08	1.50			11.44	1.50			28.72	1.57		
UNDABS	0.47	.02	.40	.13	0.47	.02	.40	.12	0.45	.02	.39	.11	0.26	.02	.22	.03
CATGRS	1.03	.10	.17	.02	1.06	.10	.17	.02	1.04	.10	.17	.02	0.74	.09	.12	.01
ANALOG	1.94	.15	.22	.04	1.96	.15	.22	.04	1.92	.15	.21	.04	1.17	.14	.13	.01
Sex: Females					-3.04	.47	-.10	.01	-3.00	.46	-.09	.01	-3.86	.43	-.12	.01
PLE: High									2.25	.48	.07	.00	-0.22†	.45	-.01	.00
Track: Medium													8.71	.59	.25	.04
Track: High													15.50	.67	.46	.10
R		.62				.63				.63				.70		
R^2 / Adj. R^2		.39				.40				.40				.50		
Unique variance		.18				.19				.18				.20		
Shared variance		.20				.20				.22				.29		

Note. All coefficients significant at $p < .001$ except †$p > .05$.

Table 7.4

Multiple regression for mathematical performance: case-control analysis

	Model 1				Model 2				Model 3				Model 4				Model 5			
	B	*SE*	β	sr^2	*B*	*SE*	β	sr^2	*B*	*SE*	β	sr^2	*B*	*SE*	β	sr^2	*B*	*SE*	β	sr^2
Intercept	7.77	1.84			8.79	1.88			10.98	1.88			11.10	1.87			25.54	1.97		
UNDABS	0.49	.03	.43	.14	0.49	.03	.44	.15	0.48	.03	.42	.14	0.46	.03	.41	.13	0.30	.03	.27	.04
CATGRS	0.81	.14	.13	.01	0.80	.14	.13	.01	0.90	.14	.14	.02	0.89	.14	.14	.02	0.71	.13	.12	.01
ANALOG	1.93	.21	.21	.03	1.91	.21	.21	.03	1.92	.20	.21	.03	1.88	.20	.21	.03	1.16	.19	.13	.01
Delay					-1.74*	.68	-.05	.00	-1.73*	.67	-.05	.00	-1.74**	.67	-.05	.00	-2.00**	.62	-.06	.00
Sex: Females									-4.80	.68	-.14	.02	-4.77	.68	-.14	.02	-5.68	.63	-.16	.03
PLE: High													2.32	.68	.07	.00	0.61†	.64	.02	.00
Track: Medium																	10.48	.83	.27	.05
Track: High																	15.29	1.07	.32	.07
R	.63				.63				.64				.65				.71			
R^2/ Adj. R^2	.39				.40				.41				.42				.50			
Unique variance	.19				.19				.21				.20				.21			
Shared variance	.20				.20				.21				.22				.29			

Note. All coefficients significant at $p < .001$ except $^{**}p < .01$; $^{*}p = .01$; $^{\dagger}p > .05$.

Case-control analysis

There were between-groups differences on the main variables for *sex* (Wilks' $\lambda = .96$, $F(4,1535) = 16.18$, $p < .001$, $\eta_p^2 = .04$) and *track* (Wilks' $\lambda = .58$, $F(8,3070) = 120.36$, $p < .001$, $\eta_p^2 = .24$). Males scored higher than females on MATH ($F(1,1538) = 43.98$, $p < .001$, $\eta_p^2 = .03$) and females scored higher than males on CATGRS ($F(1,1538) = 6.80$, $p = .009$, $\eta_p^2 = .00$) but there were no sex differences on UNDABS or ANALOG. As in the age-typical sample, scores on all four variables were highest in the high track and lowest in the low track (MATH: $F(2,1538) = 404.99$, $p < .001$, $\eta_p^2 = .34$; UNDABS: $F(2,1538) = 288.36$, $p < .001$, $\eta_p^2 = .27$; CATGRS: $F(2,1538) = 86.03$, $p < .001$, $\eta_p^2 = .10$; ANALOG: $F(2,1538) = 158.62$, $p < .001$, $\eta_p^2 = .17$; all $p_{Bonf} < .001$). The omnibus test for *delay* was not significant as there were no differences on UNDABS, CATGRS or ANALOG, but the univariate tests indicated that age-delayed students had lower MATH scores than age-typical controls ($F(1,1538) = 6.89$, $p = .009$, $\eta_p^2 = .00$). There was no effect of *PLE* and there were no 2- or 3-way interactions.

The results of the regression analyses showed similar patterns to those of the age-typical sample (see Table 7.4). The effect of delay in educational progression was small: MATH scores of age-delayed students were around 1¾ points lower than that of age-typical controls, but this did not account for any additional variance. The effect of sex was larger than in the age-typical sample, explaining 2-3% of unique variance. Females scored just under 5 points lower than males when educational track was not taken into account and more than 5½ points lower when it was. Again, the advantage of a high PLE disappeared when track was included. Post-hoc Chi-square and Fisher's exact tests found that sex showed a strong tendency to differ across tracks ($\chi^2(2) = 8.55$, $p = .014$), with proportionally fewer boys in the high track. PLE ($\chi^2(2) = 102.51$, $p < .001$) showed the same patterns as in the age-typical sample.

DISCUSSION

The goals of this large-scale, nationally representative study involving 3,663 ninth grade students in the Netherlands were twofold. First, to elucidate patterns in the relation between abstract and relational thinking skills and mathematical performance that could support attention to these skills in secondary school. Second, to provide insights into the influence of specific individual and demographic factors - namely sex, socio-economic background, educational track and delay in educational progression. Both issues are relevant for and of interest to educators and educational policymakers, given current concern to improve students' mathematical performance and to understand determinants of educational progression in lower secondary school. The large size and relative homogeneity of the age-typical sample combined with the case-control design for age-delayed students allows strong inferences to be drawn about the effects found.

The results confirm the importance of abstract and relational thinking skills for mathematical performance in the age group studied. Together, they accounted for 18% unique variance and explained 39% of the total variance in mathematical performance. The ability to understand abstract concepts was most influential. It has been argued that developmental differences in the acquisition and comprehension of abstract vocabulary during adolescence (Benelli, Belacchi, Gini, & Lucangeli, 2006; Nippold, Hegel, & Sohlberg, 1999) result in large individual differences in the ability to process educational material that includes abstract terms (Reed, Van Wesel, Ouwehand, & Jolles, 2014). Consequently, students who are better able to understand such levels of abstraction (e.g., 'differential', 'exponential', 'permutation', etc.) would be at an advantage in dealing with the corresponding subject matter. This suggests that students may benefit from specific instructional attention to understanding the meanings of these concepts before being expected to master their use. This is an important point, given the debate in mathematics education on whether students should first be taught conceptual or procedural knowledge (Schoenfeld, 2004).

There are also significant individual differences in these abilities that impact performance. A substantial proportion of these differences is attributable to *educational track*. When track is taken into account, the amount of unique variance explained by abstract and relational thinking skills is reduced to 5%. In the Netherlands, different tracks offer different curricula and instruction directed towards the type of further education or employment that students are destined for. Tracks thus differ with regard to content, coverage and level of intellectual complexity both required and aimed for. Students in tracks that prepare for vocational education are unlikely to encounter topics as matrices and polynomial functions that require them to understand quite complex abstract structures and relations. Reduced exposure to such material may in turn hold back the development of skills required to deal with it. Thus, differential educational experiences of students in different tracks may accentuate or even create and then perpetuate differences in skill levels. If this self-reinforcing pattern is to be broken, measures need to be taken to help students in all tracks to develop these skills. We return to this point later.

Although *sex* and *socio-economic background* (operationalised as the highest level of parental education (PLE)) individually predict mathematical performance, these effects are small and have limited impact on the results. It is notable that socio-economic differences are overshadowed by educational track. This appears to be attributable to an overrepresentation of students with medium PLE in the lowest track and an underrepresentation of these students in higher tracks. This is in line with research showing that early tracking can perpetuate socio-economic divisions (Hanushek & Wößmann, 2006; Marks et al., 2006; Schütz et al., 2008; Van de Werfhorst & Mijs, 2010). It should be noted, however, that students with low PLE were not included in the study due to their underrepresentation in the sample. Thus, the findings cannot be generalised to those students.

Educational delay has little effect on the relation between abstract and relational thinking and mathematical performance. Age-delayed students have slightly lower mathematics scores than age-typical controls, but do not differ with respect to abstract and relational thinking skills. Thus, although delay (in the form of grade retention) is often reported to be detrimental to student outcomes (Brophy, 2006; Jacob & Lefgren, 2009; OECD, 2012b), the findings of the present study permit an alternative hypothesis. Developmental variability means that some students lag behind their peers in the development of important thinking skills, and these students may benefit from additional time to 'come up' to the level required to deal with curricular demands. If this hypothesis holds - and future longitudinal research exploring the relationship between abstract and relational thinking and mathematical performance in several age groups should investigate this possibility - it argues strongly for timely diagnosis and intervention to help students who are at risk of delay to develop such thinking skills.

There are other reasons for intervention to help students develop abstract and relational thinking skills. For one, as argued, students in lower tracks may not be offered the educational experiences necessary to develop these skills. Also, if the previously noted lack of progress in mathematical competence during the initial years of secondary school (Ryan & Williams, 2007) is indeed due to a mismatch between curricular demands and the rate of development of these skills, it would be of great interest to investigate whether training these skills during early adolescence could improve those skills and - importantly - transfer to mathematical performance.

There are indications that helping students develop such skills is related to higher mathematical performance. Analysis of TIMSS (Trends in International Mathematics and Science Study) video data from mathematics classrooms in eight countries shows that high performance is correlated with instructional practices that give students opportunities to draw connections between mathematical ideas and reason analogically (Richland et al., 2012). Furthermore, it is not the case that these kinds of skills improve *without* intervention. Decades ago, Van Hiele (1986) claimed that higher levels of abstract thinking develop as a product of experience and instruction, a view that continues to be held by contemporary researchers. Watson (2010), for example, argues that difficulties in learning secondary school mathematics - including adopting new ways of categorisation, using abstract symbolic notations, and understanding internal relationships of mathematical objects - require changes in thinking that are unlikely to occur spontaneously and only with "particular intervention" (p. 137). Similarly, Richland et al. (2012) posit that fostering students' reasoning skills, particularly relational thinking, may be crucial to enhancing their ability to develop usable, flexible and transferable mathematics knowledge.

The development of abstract and relational thinking is also strongly related to the development of other cognitive functions, namely working memory, the ability to inhibit irrelevant information, and the ability to shift between tasks. Analogical reasoning, for example, requires source and target analogues to be mapped, evaluated and re-represented in working memory, which imposes heavy cognitive load on its limited resources (Gentner & Colhoun, 2010; Richland et al., 2012). Relational correspondences may also compete with feature-based similarities between objects, requiring inhibitory control to suppress superficial responses (Huang-Pollock, Maddox, & Karalunas, 2011; Morrison, Doumas, & Richland, 2011). Attention to superficial features could also be due to a persistence with strategies that have served well in the past, hindering shifting to another approach (Ryan & Williams, 2007; Watson, 2010). It is therefore also important to investigate the effects of training these kinds of functions.

Finally, certain aspects of the study should be addressed or extended in future research. The effect of a low socio-economic background could not be investigated, due to underrepresentation in the sample. Also, socio-economic background was operationalised as PLE; future studies could include other important indicators such as parental occupational status and home resources (see OECD, 2013b). A single test was used to measure relational thinking and the reliability of the resulting scale was just satisfactory for this type of research. Future research could investigate relational thinking with other tasks. For example, as visuospatial skills are known to be important for mathematical thinking, tests could be extended to include tasks such as progressive matrices or identifying relations within pictorial stimuli.

Acknowledgements

The authors thank the creators of the COOL$^{5-18}$ (2008) dataset, particularly Greetje van der Werf and Hans Kuypers from GION, Institute for Educational Research, University of Groningen, for the use of the dataset for this study. The COOL$^{5-18}$ dataset was obtained from the Data Archiving Network Services (DANS) website (http://www.dans.knaw.nl) with the following reference:

- GION RU Groningen, CITO Arnhem, SCO-Kohnstamm Instituut Amsterdam, ITS Radboud Universiteit Nijmegen
- Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Voortgezet Onderwijs - 2008, Eerste meting, Voortgezet onderwijs, 2008 (2007/2008)
- Persistent Identifier: urn:nbn:nl:ui:13-rrd-jk9