The Relationships among Between-class Ability Grouping, Teaching Practices, and Mathematics Achievement: A Large-scale Empirical Analysis

Cheng Yong Tan* and Clive Dimmock

Cheng Yong TAN (PhD) is Associate Professor in the Faculty of Education, The University of Hong Kong. His research interests comprise family and school influences on student academic achievement, and how familial factors moderate the effects of school processes on student achievement. Contact details: Faculty of Education, Meng Wah Complex Room 519, The University of Hong Kong, Pokfulam, Hong Kong; Tel: (852) 3917 4635; Email: tancy@hku.hk

Clive DIMMOCK (PhD) is Professor in Professional Learning and Leadership in the School of Education, University of Glasgow. His research interests are learning-centred leadership; connectivity between learning-teaching, professional development/leadership, and whole school re-design; cross-cultural comparisons between Anglo-American and Asian systems of schooling; the roles of policymakers and school leaders in promoting equity across school systems; and problems arising from disconnections between research, policy and practice. Contact details: School of Education, St Andrew's Building, 11 Eldon St, University of Glasgow, G3 6NH Glasgow, UK; Tel: +44 (0)141 330 3037; Email: Clive.Dimmock@glasgow.ac.uk

Running Head: ABILITY GROUPING AND MATHEMATICS ACHIEVEMENT

Word Count: 8,595

---

* Corresponding author

The Relationships among Between-class Ability Grouping, Teaching Practices, and Mathematics Achievement: A Large-scale Empirical Analysis

Abstract

The knowledge base on various forms of structuring students' learning by ability grouping is more robust than that on teachers' instructional practices being implemented within these groupings. The present study examines if student-reported mathematics teachers' instructional practices vary among schools with different degrees of implementing between-class ability grouping and if between-class ability grouping moderates the relationship between these practices and students' mathematics achievement. International data from 281,591 fifteen-year-old students and 11,765 principals were analyzed. One-way ANOVA showed that students perceived greater implementation of teachers' instructional practices in schools with more between-class ability grouping. However, hierarchical linear modeling (HLM) results showed that between-class ability grouping did not moderate the relationship between teachers' instructional practices and students' mathematics achievement. These results imply that it is more important to examine the relationship between teachers' instructional practices and students' achievement in homogenous ability classes vis-à-vis that between ability grouping and students' achievement *per se*.

Keywords: ability grouping; teachers' instructional practices; mathematics achievement; PISA; secondary schools

Introduction

Schools have different ways of organizing students for effective teaching-and-learning. Some provide common instruction and curriculum for every student (mixed grouping) whereas others differentiate these aspects for students (ability grouping), presumably based on students' initial readiness, needs, and abilities. The debate on whether to group students in different classes for different subjects, including mathematics, in schools has ensued for decades (Chmielewski 2014; Gamoran 2010; Gentry 2016; Johnston and Wildy 2016; Schofield 2010). The search for effective instructional strategies to teach mathematics reflects the importance of mathematics and science mastery for participating in science-related higher education and jobs (Xie, Fang, and Shauman 2015). However, many schools struggle in identifying the most effective organizational configurations and instructional strategies to enhance students' attitudes toward learning these subjects that contribute to their achievement (Osborne, Simon, and Collins 2003).

Schools may adopt one or more of three main types of ability grouping, namely the assignment of students to different schools (between-school grouping), different classes within a school (between-class grouping), and/or different groups within a class (within-class grouping) (Chmielewski 2014; Chmielewski, Dumont, and Trautwein 2013; Huang and Hsieh 2019; Johnston and Wildy 2016; Steenbergen-Hu, Makel, and Olszewski-Kubilius 2016). In particular, between-class ability grouping in secondary mathematics is prevalent in many education systems worldwide (Chmielewski et al 2013; Johnston and Wildy 2016; Wilkinson and Penney 2014). Between-class ability grouping comprises two common arrangements (Chmielewski et al 2013). First, students can be assigned to different tracks, programmes, or streams for all subjects within the school (e.g., academic, general, and vocational tracks) as is the case in some schools in Belgium, France, Germany, Ireland, Luxembourg, the Netherlands, Portugal, and Switzerland.

Second, students may be tracked for specific subjects only within the school (e.g., mathematics courses at basic, regular, and advanced levels) as is the case in some schools in Australia, Germany, Iceland, and the US.

Despite its prevalence in many education systems, there is evidence that between-class ability grouping may be less effective than other forms of ability grouping. For example, Steenbergen-Hu and colleagues' (2016) comparison of meta-analytic results failed to find evidence that student learning benefited from between-class ability grouping as much as they did from other forms of ability grouping (within-class ability grouping, cross-grade subject grouping, special grouping for the gifted). However, the study detected a small positive effect of between-class ability grouping for middle and junior high school students' academic achievement. Therefore, it is unclear from the study if between-class ability grouping was positively associated with secondary students' mathematics achievement. More importantly, there is a paucity of research that examines whether teachers have in the first place, taken advantage of ability grouping to adapt their instructional practices, and if they have, whether specific instructional practices are more effective in homogeneous ability classes. The present study addresses these knowledge gaps by investigating relationships among between-class ability grouping, instructional practices, and students' mathematics achievement.

Literature review

The argument for more effective teaching with ability grouping

Creemers and Kyriakides' (2008) "dynamic model of educational effectiveness" suggests that different contextual factors (student, classroom, school, sociocultural) operate in concert, with higher-level factors facilitating or limiting the potential of lower-level factors to influence educational effectiveness. Between-class ability grouping is a school-level configuration that enables students to be sorted into different classes, presumably according to their learning abilities, so that teachers can teach more effectively.

The most compelling argument for ability grouping is perhaps the potential for teachers *in theory* to adapt their instructional strategies to address students' learning needs and improve students' learning in homogeneous ability classes (Gamoran 2010; Gentry 2016; Johnston and Wildy 2016; Schofield 2010). Indeed, teachers may prefer to teach classes where students are more similar in terms of their needs and abilities (Forgasz 2010; Markow and Cooper 2008; Tieso 2003). For example, Forgasz's (2010) study of a small sample of Years 7 to 10 mathematics teachers in Australia found that most of them supported ability grouping as they believed that it enabled high-ability students to learn at a more competitive pace, and low-ability students to gain confidence in learning mathematics through a less demanding curriculum. Indeed, the literature seems to assume that teachers will inexorably differentiate their instruction to cater to students' specific needs and abilities in homogeneous groups and benefit students' learning (Gamoran 2010; Gentry 2016; Johnston and Wildy 2016; Schofield 2010).

Mathematics teachers' instructional practices

Some teachers of homogeneous ability classes rely on a core set of instructional practices targeting a particular group of students (Boaler, Wiliam, and Brown 2000; Domina, Hanselman, Hwang, and McEachin 2016). For example, Boaler and colleagues' (2000) study of secondary students which moved from heterogeneous to homogeneous ability grouping for mathematics in the UK found that teachers of homogenous-ability classes employed more focused instructional approaches to cater to their students' abilities than peers of heterogeneous-ability classes who had to use various pedagogies to address diverse students' needs. Teachers of homogeneous-ability classes may also anticipate a more similar pattern of responses to their instruction from students than they would from students in heterogeneous-ability classes. This is not to say that all students in homogeneous-ability classes are identical in their needs and responses (Boaler et al 2000); rather, there may be less diversity in students' needs and anticipated responses by virtue of the ability grouping process. With the smaller variation in students' learning abilities, teachers can teach a more demanding curriculum to higher-ability classes and vice versa (Domina et al. 2016). For example, Domina and colleagues' (2016) study of the implementation of the algebra-for-all initiative in California found that despite accountability pressures to offer algebra to all eighth graders, high-socioeconomic status (SES) middle schools continued to differentiate their curriculum and cater to their students' abilities, offering advanced geometry courses to high-ability students and algebra courses to low-ability students.

Mathematics teachers of homogenous- and heterogeneous-ability classes may use different instructional practices. A review of the teacher effectiveness literature indicates that there is a set of instructional practices that have been demonstrated to address students' needs and benefit their learning (Baumert et al. 2010; Creemers and Kyriakides 2008; Day et al 2008; Klieme, Pauli, and Reusser 2009; Reynolds et al 2014). These practices are cognitive activation, classroom management, and personal learning support.

Cognitive activation

Cognitive activation refers to teachers providing students with opportunities to solve challenging problems and helping students to develop and apply strategies to solve authentic problems (Baumert et al. 2010; Klieme et al 2009). Cognitive activation builds on students' prior knowledge and enables teachers and students to co-construct knowledge. In so doing, students experience ownership, empowerment, and engagement in their learning (Day et al 2008). There is evidence that cognitive activation is commonly practised in mathematics classrooms (Chirinda and Barmby, 2018; Conner, Singletary, Smith, Wagner, and Francisco 2014; Ellis, Özgür, and Reiten 2019; Larsson and Ryve 2012). For example, secondary mathematics teachers can directly ask questions related to argumentation or provide support for argumentation (Conner et al 2014). They can also cognitively engage students by using different types of routines (practical/discursive, process-/product-oriented; Lavie, Steiner, and Sfard 2019).

Mathematics teachers of homogenous-ability classes may find it easier to teach problem-solving skills to their students (Carbonaro 2005; Forgasz 2010). To illustrate, Forgasz (2010) reported that her sample of Years 7 to 10 mathematics teachers in Australia differentiated their teaching among classes of different ability levels, with those teaching high-ability classes focusing on higher-level conceptual understanding and open-ended assignments, authentic learning activities, and even extended enrichment activities, and those teaching low-ability classes addressing basic skills development in their students such as distinguishing relevant from irrelevant information in questions and training students to answer each part of a question sequentially. In another study, Carbonaro's (2005) analysis of 8^{th}-10^{th} graders found that students in high-ability classes experienced a more demanding mathematics curriculum and intellectually stimulating instruction. They then responded to higher learning expectations by putting in more effort in their learning which benefited their mathematics achievement.

The preceding discussion assumes that if students are assigned to the "correct" classes based on their abilities, mathematics teachers will be able to provide problem-solving learning experiences that commensurate with students' cognitive abilities. However, some critics contend that the assignment process may not be based on academic merit (Dunne et al. 2007; Forgasz 2010; Paul 2005). For example, Dunne and colleagues' (2007) analysis of 44 secondary students in the UK found that students' prior abilities did not necessarily predict their set placement. For example, only 46.9% of low-ability students, 53.2% of average-ability students, and 56.5% of high-ability students were correctly assigned to the correct mathematics sets. Other criteria besides academic ability that predicted students' set placement were students' residential neighbourhoods, free school meal eligibility, special educational needs, and ethnicity.

Classroom management

Classroom management refers to teachers setting clear routines and managing disruptions to maximize quality learning time for students (Baumert et al. 2010; Klieme et al. 2009). Classroom management is important in the teaching of different subjects, including mathematics (Day et al, 2008; Juta and Van Wyk 2020). For example, mathematics teachers can organize student groupwork and motivate students in their classroom management (Juta and Van Wyk 2020). Day and colleagues' (2008) study also found that Year 9 students in UK schools perceived that they had learnt effectively in their mathematics classes if their teachers could, among other things, provide a conducive classroom environment (less misbehaviours) for them to concentrate on learning.

Mathematics teachers of homogeneous-ability classes may find it easier to implement classroom management practices to address students' needs. For instance, they may use a combination of intrinsic and extrinsic reinforcement for low-ability classes and more intrinsic motivational strategies for high-ability classes (Cheng, McInerney, and Mok 2014). However, Hallam and Ireson's (2005) study of mathematics, English, and science teachers in secondary comprehensive schools in the UK found that teachers of heterogeneous-ability classes spent less time getting low-ability students to behave compared to peers teaching homogeneous low-ability classes. What is not apparent from this study is whether mathematics teachers particularly experienced these classroom management challenges.

Personal learning support

Personal learning support refers to teachers adapting instructional strategies to support individual students' learning (Baumert et al. 2010; Klieme et al. 2009). This support is important as mathematics students appreciate teachers who understand their learning needs and styles, and who can help them to build on their prior knowledge in their learning (Day et al 2008). Students benefit from teachers' support in their learning of different subjects, including mathematics (Marais, Van der Westhuizen, and Tillema 2013; Nicol 2002; Prediger, Fischer, Selter, and Schöber 2019).

Indeed, students from low- and high-ability classes may differ with regards to their learning needs. This is evident in Zevenbergen's (2003, 2005) study of Years 9 and 10 students' mathematics learning in Australia which found that students in high-ability classes identified themselves with mathematics, were positive about mathematics, were confident of their mathematics achievement, and expected to continue studying mathematics into Year 11. However, students in low-ability classes conceded that they were weak in mathematics, disliked mathematics, felt inferior about not studying mathematics, and did not intend to study mathematics in future. With ability grouping, mathematics teachers may find it easier to cater to students' needs. For example, they can break down complex ideas into smaller parts and underline how these parts constitute the whole (decompressing) or intentionally omit details (trimming) for low-ability classes, and connect different content areas (bridging) or add details to enrich students' understanding (trimming) for high-ability classes (McCrory, Floden, Ferrini-Mundy, Reckase, and Senk 2012).

Teachers can also support student learning by using formative assessment to identify students' learning needs, provide feedback to students, and evaluate and adapt their own instruction (Creemers and Kyriakides 2008). Feedback can benefit students' learning by informing learners on their learning gaps and the development of a roadmap to reach learning goals (Hattie and Timperley 2007). Given these potential learning benefits, teachers of different subjects, including mathematics (Chen, Crockett, Namikawa, Zilimu, and Lee 2012), are encouraged to use formative assessment in their teaching.

Mathematics teachers may find it easier to provide formative feedback for homogeneous- (vis-à-vis heterogeneous-) ability classes. For example, Hallam and Ireson's (2005) UK study found that, compared to peers teaching heterogeneous-ability classes, teachers teaching different subjects (including mathematics) for homogeneous-ability classes reported that they were more proficient in providing detailed written feedback on homework for high-ability students and setting less homework for low-ability students.

Present study

The specific research objectives of the present study are to -

- examine if teachers' instructional practices vary among schools with different degrees of implementation of between-class ability grouping in mathematics; and
- clarify if the association between teaching practices and students' mathematics achievement is moderated by the implementation of between-class ability grouping.

The study follows the methodological design in many studies of ability grouping in using Programme for International Student Assessment (PISA) data to examine ability grouping or variables related to ability grouping across different education systems (e.g., Chmielewski 2014; Chmielewski et al 2013; Mijs 2016; Robert 2010). This approach enables researchers to leverage large-scale international data that are otherwise difficult to collect. However, the data are limited in that they do not capture all variants of ability grouping. For example, some education systems divide secondary education into lower and upper secondary levels, and entry into upper secondary tracks depends on some form of ability grouping. Therefore, upper secondary levels in these education systems may not have further between-class ability grouping arrangements in schools. Having said that, the analysis in the present study showed that 29 of the 34 Organization for Economic Cooperation and Development (OECD) countries examined had schools that employed between-class ability grouping for most of their students; the other five countries that did not (Austria, Czech Republic, Greece, Norway, Slovenia) still had 16.93-46.32% of schools that implemented ability grouping for some or all classes (Supplementary Table 1).

The present study used students', as opposed to teachers', reports of instructional quality. There are methodological strengths associated with each type of data (Kunter and Baumert 2006). Specifically, Kunter and Baumert (2006) found that both types of data demonstrated construct validity in measuring aspects of instructional quality related to students' cognitive autonomy and teachers' classroom management. There was substantial agreement between students' and teachers' ratings on classroom management. Students' ratings were particularly useful in measuring teachers' support whereas teachers' ratings provided useful information on the use of instructional tasks and methods. There are three covariates in the analysis (student

SES, gender, whether the country has a Confucian heritage culture (CHC)) because studies showed that these variables were associated with levels of students' academic performance (Else-Quest, Hyde, and Linn 2010; Tan 2020; Tan and Liu 2017).

Supplementary Table 1

Method

Participants

Participants were 281,591 students and their principals from 11,765 schools in 34 OECD countries who participated in the PISA 2012 (OECD 2013). These countries had varying numbers of students who attended schools with students' ability grouping for some or all classes in mathematics instruction, from 16.93% in Greece[1] to 99.28% in Ireland[2] (Supplementary Table 1). 41.99% in the entire sample attended schools with ability grouping for some mathematics while another 35.38% attended schools with ability grouping for all mathematics classes.

Participating students in PISA 2012 were selected to represent the complete population of 15-year-old students who were attending public or private schools in grade 7 or higher in the participating countries. PISA 2012 measured 15-year-old students' proficiency in applying their knowledge and skills learned in mathematics (the focal domain) in addition to reading and science. In addition, PISA 2012 also collected background data from students and principals on students/home/family, classroom, and school variables. Students in the sample consisted of a

[1] (535 + 323) / 5,067 * 100

[2] (1,868 + 2,693) / 4,594 * 100

balanced mix of boys and girls (i.e., 50% each). Their mean SES score was -0.06 (SD = 1.05; range = -5.62 to 3.27). Two out of the 34 countries were CHCs (Japan, Korea).

Measures

Data on the following variables from the PISA 2012 dataset were used in the analysis.

First, students' mathematics achievement was the focal achievement variable measured in PISA 2012. Students were not administered the complete set of test items by design, and therefore each item had missing responses. This made it impossible to estimate achievement scores for each student. To overcome this limitation, PISA aggregated the results of individual students to produce scores for groups of students and used students' responses to the assessment and survey background items to compute a set of five "plausible values" (PVs) for each student (MathPV1 to MathPV5) for estimating students' mathematics scores.

Principals responded to two items on the extent to which students were grouped by ability into different classes either with similar context but different difficulty levels or with different content taught within the school. PISA 2012 then derived an index (ABGMATH) measuring the extent to which schools implemented ability grouping between mathematics classes (1=*No ability grouping*; 2 = *Ability grouping for some classes*; 3 = *Ability grouping for all classes*). For the purposes of the present study, this index was recoded as SomeClasses (=1 if ABGMATH=2, 0 otherwise) and AllClasses (=1 if ABGMATH=3, 0 otherwise).

PISA 2012 derived Rasch measures measuring mathematics teachers' instructional practices based on students' responses to four sets of items (OECD 2014). These measures were:

- COGACT: Nine items (e.g., "The teacher presents problems for which there is no immediately obvious method of solution") measuring the frequency to which teachers implemented cognitive activation in mathematics classes (1 = *Always or almost always*; 2 = *Often*; 3 = *Sometimes*; 4 = *Never or rarely*; reverse coded);

- CLSMAN: Four items (e.g., "My teacher keeps the class orderly") measuring the extent to which mathematics teachers effectively managed their classes (1 = *Strongly agree*; 2 = *Agree*; 3 = *Disagree*; 4 = *Strongly disagree*; reverse coded);
- TCHBEHFA: Four items (e.g., "The teacher gives me feedback on my strengths and weaknesses in mathematics") measuring the frequency to which mathematics teachers provided personal learning support using formative assessment (1 = *Every lesson*; 2 = *Most lessons*; 3 = *Some lessons*; 4 = *Never or hardly ever*; reverse coded); and
- TEACHSUP: Five items (e.g., "The teacher gives extra help when students need it") measuring the frequency to which mathematics teachers provided personal learning support by addressing students' learning needs (1 = *Every lesson*; 2 = *Most lessons*; 3 = *Some lessons*; 4 = *Never or hardly ever*; reverse coded).

These four instructional variables had been used to measure cognitive activation, classroom management, formative assessment, and teacher support in other studies analyzing PISA data (Tourón, Navarro-Asencio, Lizasoain, López-González, and Pedro 2019; Yi and Lee 2017).

Three control variables were included:

- ESCS: Students' SES was measured by a PISA variable summarizing their parents' educational attainment, family wealth, and occupational status;
- Male: Students' sex was measured using a dummy variable (0 = girls, 1 = boys);

- CHC: This variable was coded as 1 for countries which were Confucian heritage cultures and 0 otherwise.

Procedure

PISA 2012 used a two-stage stratified sampling design, with schools first selected from a national sampling frame of schools with probabilities proportional to size and students then selected from within each selected school (Liou and Hung 2015). PISA 2012 was sponsored internationally by the OECD and coordinated and administered internationally by the PISA international consortium, led by the Australian Council for Educational Research. All participating economies followed standardized procedures outlined in the technical standards and manuals provided. PISA 2012 data are publicly available (no prior permission needed) for research purposes (data source: https://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm).

Data analysis

There were missing values (0 - 35.13%) in the data which might compromise estimation efficiency and produce biased results. Therefore, Markov chain Monte Carlo multiple imputation was employed to address the methodological challenge arising from missing values in the data. This multiple imputation procedure is a generally more effective method of data imputation compared with other procedures for handling missing values, and especially useful for large samples or data with higher percentages of missing values (Cheema 2014). The multiple imputation procedure imputed missing values five times, thereby producing multiple complete datasets. In the present study, the pooled parameter estimates (across the five imputed datasets) are analyzed. The standard errors of the estimates are unbiased when this procedure is used.

Three-level fixed effect HLM with maximum likelihood estimation (Raudenbush and Bryk, 2002) was performed using HLM7 (Raudenbush, Bryk, Cheong, Congdon, and Du Tolt 2011) to examine relationships among between-class ability grouping, teachers' instructional practices, and students' mathematics achievement. HLM was used to account for the nested structure of the data, test for mean differences in dependent variables (students' achievement), incorporate both continuous and categorical variables, account for the different sizes of units, and compute the proportion of explained variance at different levels (Dedrick et al. 2009). HLM7 first estimated parameters for each of the five PVs (PISA-computed) before averaging the estimates in the HLM analysis. It then combined the average of the sampling error from the five PVs with the variance between the five PVs multiplied with a factor related to the number of PVs.

In the analysis, the independent variables were rescaled by subtracting the grand mean of the entire sample from the respective raw scores for ease of interpretation. After the rescaling, each HLM parameter represents the "effect" of the respective variable for a student with values on the other variables that are each equal to the grand mean for the respective variable. The following set of HLM models was fitted:

- Model 1 - baseline with no predictors;
- Model 2 – random intercepts model with control variables;
- Model 3 – random intercepts model with control variables and between-class ability grouping variables;

- Model 4 - random intercepts model with the variables in Model 3 and teachers' instructional practice variables; and
- Model 5 – random intercepts model with the variables in Model 4 and interaction variables involving between-class ability grouping and teachers' instructional variables.

Results

Did teachers' instructional practices vary among different schools?

One-way ANOVA and post-hoc Tamhane test were used to compare levels of teachers' instructional practices among schools with no ability grouping for mathematics, ability grouping for some mathematics classes, and ability grouping for all mathematics classes (Supplementary Table 2). Results showed overall significant differences for the perceived implementation of teachers' instructional practices among the three types of schools were significant, $p < .001$.[3] In particular, compared to peers in schools with no or some classes grouped by students' ability, mathematics teachers in schools where all classes were grouped by students' ability were perceived to have implemented cognitive activation (COGACT, $F[2, 183464] = 110.20$), classroom management (CLSMAN, $F[2, 182771] = 94.54$), formative assessment (TCHBEHFA, $F[2, 183500] = 128.10$), and addressed students' learning needs to a greater extent (TEACHSUP, $F[2, 183992] = 378.94$).

Supplementary Table 2

---

[3] All post-hoc pairwise comparisons were significant at the .05 level.

Did ability grouping moderate relationships between teachers' instructional practices and students' mathematics achievement?

HLM Model 1 results (Table 1) showed that there was variance in students' mathematics achievement occurring within schools at level 1 (52.69%), between schools at level 2 (37.15%), and between countries at level 3 (10.17%), thereby affirming the utility of using HLM for the present analysis. Model 2 results showed that boys (Male; $\pi = 12.37$), high-SES students (ESCS; $\pi = 19.39$), and students in CHCs ($\gamma = 47.61$) had higher levels of mathematics achievement than girls, low-SES students, and students in non-CHCs respectively, $p < .001$. Model 3 results showed that SomeClasses ($\beta = -14.01$) and AllClasses ($\beta = -22.03$) were both negatively related to students' mathematics achievement, $p < .001$. Model 4 results showed that teachers' cognitive activation (COGACT, $\pi = 3.19$, $p < .001$), classroom management (CLSMAN, $\pi = 3.99$, $p < .001$), and addressing students' learning needs (TEACHSUP, $\pi = 3.28$, $p < .01$) were positively associated with students' mathematics achievement. In contrast, formative assessment was negatively related to students' mathematics achievement (TCHBEHFA, $\pi = -8.30$, $p < .001$). Model 5 results showed that the implementation of between-class ability grouping did not moderate the relationships between students' mathematics achievement and cognitive activation (SomeClasses*COGACT, $\beta = -5.12$, $p = .16$; AllClasses*COGACT, $\beta = -7.00$, $p = .13$), classroom management (SomeClasses*CLSMAN, $\beta = 4.83$, $p = .15$; AllClasses*CLSMAN, $\beta = 1.47$, $p = .60$), formative assessment (SomeClasses*TCHBEHFA, $\beta = 5.13$, $p = .09$; AllClasses*TCHBEHFA, $\beta = 3.35$, $p = .29$), and teachers addressing students' needs (SomeClasses*TEACHSUP, $\beta = -5.10$, $p = .21$; AllClasses*TEACHSUP, $\beta = 2.00$, $p = .68$).

Table 1

Discussion

Ability grouping facilitates implementation of instructional practices

The present study provided interesting insights on relationships among between-class ability grouping, teachers' instructional practices, and students' mathematics achievement. First, ANOVA results indicate that ability grouping provides the organizational structure for mathematics teachers to implement *potentially* effective instructional practices such as cognitive activation, classroom management, and personal learning support (Schlesinger and Jentsch 2016). Students whose mathematics teachers implement cognitive activation practices can hone their high-level thinking. Students in effectively managed mathematics classrooms can participate in more high-level thinking, enjoy more time-on-task, and have greater self-determination in their learning. Students who enjoy mathematics teachers' personal learning support are better able to be intrinsically motivated to learn (i.e., self-determination). High-level thinking, time-on-task, and self-determination in turn enhance students' conceptual understanding, motivation, and achievement in mathematics (Praetorius, Pauli, Reusser, Rakoczy, and Klieme 2014).

Effectiveness of instructional practices in the context of ability grouping

However, HLM results showed that the implementation of between-class ability grouping did not moderate the relationships between teachers' instructional practices and students' mathematics achievement. These results may arise for different reasons.

Non-academic criteria in class assignment

The first reason relates to the process of assigning students to homogeneous-ability classes. This process is supposed to be based on an accurate assessment of students' prior academic abilities, but there is some evidence that it may instead be influenced by teacher inferences based on students' gender, SES, ethnic origins, and even behaviour and inferred motivation (Forgasz 2010; Paul 2005; Taylor et al. 2019). For example, Paul's (2005) analysis of urban schools found that there were more Asian-American and White than African-American and Hispanic students who were enrolled in Algebra 1, an important mathematics course for college preparation. Teachers in Forgasz's (2010) study of a small sample of Years 7 to 10 mathematics teachers in Australia conceded that students' assignment into different ability groups in their schools were not solely based on students' marks or grades. Other criteria impacting assignment included teachers' recommendations and students' and parents' choices. The influence of non-academic criteria impinging on the process of assigning students to homogenous-ability classes means that teachers may not be able to effectively implement cognitive activation, classroom management, and the provision of personal support.

Teacher support addressing 'range' of student needs even within 'homogenous' classes

Even if students are accurately assigned to classes of different abilities, there may still exist a range of ability levels within each class. Therefore, mathematics teachers need to recognize that students in homogeneous-ability classes can vary in their learning styles, learning pace, and perceptions of task difficulty (Boaler et al. 2000). Teachers' difficulty in catering to this variation in students' needs in homogeneous-ability classes means that higher-ability students may be bored when teachers teach to the mean, and lower-ability students may be frustrated and become disruptive when teachers do not offer sufficient support for their learning, thereby giving rise to classroom management problems (Ofsted 1993).

Furthermore, students' abilities in homogenous-ability classes can vary across different mathematics topics. Given the importance of catering to students' needs, teachers need to differentiate the problem-solving strategies taught, encourage students for their effort (especially low-ability students) and success in problem-solving (Carbonaro 2005), employ different strategies to keep students engaged in classroom interactions (Battista 2010), and use different motivation strategies to maintain a positive classroom learning climate (Cheng et al. 2014).

Implementing cognitive activation more effectively

It is also important to differentiate between teachers' implementation of instructional practices and the effectiveness of these practices. In the case of cognitive activation, teachers need to progress from eliciting students' answers to eliciting students' ideas, from correcting students' errors to prompting error correction, from cueing to providing guidance, and from encouraging valuation to encouraging reflection (Ellis et al 2019). Lavie and colleagues (2019) argued that it was insufficient for students to just perform mathematics learning tasks (ritualization); rather, teachers must help students to de-ritualize and progress to focusing on the product of the learning (i.e., exploration). Specifically, mathematics teachers could let students (a) realize that there was more than one way to perform the task; (b) appreciate that the outputs in any given step could be inputs for successive steps; (c) understand that their learning could generalize beyond the present task; (d) progress to setting the task to address the latter's own needs independently of others' involvement; (e) relate the task to mathematical abstraction as opposed to concrete objects; and (f) shift the latter's focus from demonstrating that they have performed the mathematical procedure correctly to the substantive outcome. Indeed, Larsson and Ryve (2012) concluded from their study of key mathematical practices in whole-class discussions in a seventh grade class that mathematics teachers might be implementing procedure- instead of competency-oriented practices because of the complexity of mathematics problems and the teaching process, thereby limiting students' development of broad mathematics competencies.

Re-examining types of feedback given in formative assessment

Similarly, there is room for teachers to improve their implementation of formative assessment. For example, Hattie and Timperley (2007) argued that teachers need to discriminate among different types of feedback provided to students. They highlighted the importance of giving feedback not just for the task at hand but also for the way students "process" learning tasks and for students' self-regulated learning, while eschewing giving feedback about students themselves. Providing formative feedback on students' task processing and self-regulation is important especially for students in low-ability classes as these students may have lower mathematics self-concepts (Chmielewski et al. 2013) and they may invest less effort in their mathematics learning (Carbonaro 2005). Students in high-ability classes will also benefit from more constructive feedback on their learning processes, given that some of these students are exposed to surface instead of deep learning orientations (Boaler 1997b). Indeed, Chen and colleagues' (2012) study of eighth grade mathematics teachers' discourse-based assessment practices found that some teachers used formative assessment to promote deep understanding whereas others used it to foster rote understanding in their students. Therefore, there is scope for mathematics teachers to improve their use of formative assessment.

Differential effectiveness for low- versus high-ability classes

Yet another reason why results showed that ability grouping did not moderate relationships between instructional practices and students' achievement is that instructional practices may be more effective for some students (e.g., high-ability classes) and less so for others (e.g., low-ability classes), so there is no overall impact. Indeed, teachers may find it harder to reach out to academically weaker students. Marais and colleagues (2013) unravelled complexities characterizing teachers' knowledge of eighth graders' help-seeking behaviour in mathematics classrooms: these students preferred to seek help from classmates than teachers, they would not seek help if they perceived that their teachers were not supportive, they did not want to feel incompetent, they felt uncomfortable asking for help, they were not engaged in their learning, or that they were not aware that they needed help.

Relating classroom management to student learning processes and difficulties

It is also possible that the effectiveness of an instructional practice depends on whether teachers manage to effectively implement other practices concurrently. In the case of classroom management, there is evidence that even with between-class ability grouping, mathematics teachers may not adequately understand how individual students learn and the challenges that students face (Boaler 1997a; Boaler et al. 2000; Forgasz 2010), thereby compromising the effectiveness of classroom management practices to facilitate student learning. For example, Forgasz (2010) reported that her sample of Years 7 to 10 mathematics teachers in Australia acknowledged that the process of assigning students into classes of different ability levels did not include an adequate understanding of how students learn and sometimes was influenced by logistical factors related to the timetable and the capacity of each classroom. In another study, Boaler's (1997a) longitudinal study of students' mathematics learning as they progressed from Year 9 to Year 11 in the UK showed that despite teachers customising their teaching according to the perceived average levels of different ability classes, students expressed pressure and anxiety in having to follow the pace of the teaching (too fast for some in high-ability classes and too slow for some in low-ability classes) and the frustration of having their learning progress adversely affected as a result of the ability grouping. Similarly, Boaler and colleagues' (2000) study of students in secondary schools which moved from heterogeneous- to homogeneous-ability grouping for mathematics in the UK, found that lower-ability classes experienced less opportunity to learn whereas higher-ability classes were expected to learn at an unreasonably fast pace. If teachers are unable to address students' learning needs, they may find it difficult to address classroom management issues arising from students' pressure, anxiety, and frustration.

Challenges of implementing cognitive activation in context

Teachers need to consider the context in which they implement instructional practices. Prediger and colleagues' (2019) quasi-experimental study examining how mathematics teachers could support low-achieving secondary students in Germany highlighted the importance of community, school, and district resources and support complementing classroom variables (e.g., having well-designed teaching materials) for student learning to improve. For example, contextual challenges such as classroom disciplinary issues and students' inadequate mathematics problem-solving skills and unfamiliarity with the language of teaching and learning (Chirinda and Barmby 2018) may impede the effectiveness of cognitive activation. Less efficacious mathematics teachers may have difficulty teaching lessons that are cognitive demanding, that extend students' explanations, that promote student-to-student discourse, and that connect abstract representations (Lee, Walkowiak, and Nietfeld 2017). Lastly, teachers may not be able to relate abstract mathematics learning to authentic work settings if they do not have applied mathematics backgrounds (Nicol 2002).

Gender, SES, and CHCs

Lastly, results on relationships between the three control variables and students' mathematics achievement merit discussion. First, the finding that boys outperformed girls in their mathematics achievement adds to the body of evidence on the continuing gender gap (in favour of boys) in students' mathematics attitudes, affect, and performance (Else-Quest et al 2010). Second, the finding that higher-SES students outperformed peers from disadvantaged families relates to their privileged access to different family resources (e.g., cultural capital) that benefit their mathematics learning (Tan 2020). The higher levels of mathematics performance of students from CHCs may be attributed to the sociocultural emphasis on learning and national investments in education in these countries (Tan and Liu 2017).

Conclusion

Theoretical contributions

The present study contributes to relevant scholarship by demonstrating that school structural arrangements (e.g., between-class ability grouping) are not deterministic in impacting students' learning. Therefore, we need to reconsider the widespread practice of assigning students to different classes for mathematics in secondary schools and instead focus on instructional processes that are occurring within classrooms (Gamoran 2010; Gentry 2016; Johnston and Wildy 2016; Schofield 2010). As Baines (2013) has averred, "it is not the activity of between-class ability grouping per se that leads to the observed small effects on achievement but rather its interaction with curriculum, classroom, and students' characteristics" (p.118). Indeed, instructional processes can either actualize or undermine the efficacy of the best-intentioned ability grouping practices.

Social justice implications

The practice of grouping students for learning in schools has been continuously debated for its ramifications for social justice (Boaler 1997a; Domina et al. 2016). For example, students assigned to low-ability classes have to contend with ceilings set for their achievement. This is evident in Boaler's (1997a) longitudinal study of students' mathematics learning in Year 9 to Year 11 in the UK showing that students in low-ability sets perceived their potential achievement in mathematics as being constrained not so much by their effort but by the set they were assigned. Indeed, students in low-ability groups risk being disadvantaged in many ways, including misallocation to groups, difficulty in being reassigned to high-ability groups in future, and being taught by poorly qualified and ineffective teachers (Francis et al. 2017). Mazenod and colleagues (2019) argued that students in low-ability classes may be too reliant on their teachers, thereby forgoing opportunities for independent learning. Furthermore, the different curricula offered to students from different ability classes constrain their future education. For example, Domina and colleagues' (2016) study of the implementation of the algebra-for-all initiative in California found that only secondary students in high-ability classes were able to take higher-level mathematics courses which then facilitated their access to advanced mathematics courses in high school.

However, proponents of "detracking" have to contend with pressures supporting and perpetuating ability-grouping (Burris, Heubert, and Levin 2006; Domina et al. 2016). For example, Domina and colleagues' (2016) analysis of the algebra-for-all period in California found that despite policymakers' pressures for secondary schools to "detrack" their mathematics curriculum, many schools continued to offer a separate more advanced track (e.g., teaching geometry) alongside the new baseline algebra track to their more able students (who often come from high-SES families). Additionally, schools face societal pressure for differentiation to cater to students' needs and abilities, political pressure from teachers who prefer to teach high-ability classes and parents who favour high-ability classes, and technical pressure from teachers of heterogeneous-ability classes (Gentry 2016; Johnston and Wildy 2016; Markow and Cooper 2008; Schofield 2010; Tieso 2003).

To resolve this quandary, students should first be assigned to homogeneous-ability classes based on a more comprehensive set of objective performance criteria. This more informed class assignment will enable teachers to better address students' learning needs, manage classroom issues, and provide individualized, context-specific support to students.

Limitations

There are three limitations associated with the study. First, it focuses on the implementation of between-class ability grouping but in practice, students may also be contemporaneously subjected to within-class ability grouping in some schools (Chmielewski 2014; Chmielewski et al. 2013; Johnston and Wildy 2016; Steenbergen-Hu et al 2016). Furthermore, there can be huge variation among the category of schools implementing between-class ability-grouping for some classes in the data. For example, some of these schools may group high-ability students into one or two top classes and put the remaining students into mixed-ability classes. Other schools may form homogeneous-ability classes for high- and low-ability students and heterogeneous-ability classes for average-ability students. Schools may also group students by ability for some difficult mathematics topics but not others. Lastly, some education systems stratify junior secondary students by ability for entry into senior secondary levels, so there may not be further between-class grouping at senior secondary levels. PISA data do not provide information on these different grouping variations, so results reported in this study should be read with these caveats in mind.

Second, the study relied on student-reported data on teachers' instructional practices, and students' perceptions may be limited in their assessment of some aspects of teachers' instructional practices as compared to classroom observations by experienced external assessors who are able to make less subjective evaluations of instructional quality (Schlesinger and Jentsch 2016). The third limitation is that the PISA 2012 data are cross-sectional, so the present study can at best allude to, but not definitively establish, relationships among between-class ability grouping, teachers' instructional practices, and students' mathematics achievement.

Future research

Future research is needed that extends the investigation to examine the diversity and effectiveness of teachers' instructional practices, using alternative sources of data on teachers' instructional practices such as lesson observations by external experts (Schlesinger and Jentsch 2016), in the context of myriad ability grouping arrangements (e.g., between-school, within school, and within-class ability grouping). Further studies are also needed to explore how teachers' instructional practices and students' peer interactions interact to affect students' learning. Only then can a more complete picture emerge of the complex relationship among ability grouping, teachers' instructional practices, and students' mathematics achievement.

References

Baines, E. 2013. “Ability Grouping.” In *International Guide to Student Achievement,* edited by J. Hattie and E. M. Anderman, 116-118. New York: Routledge.

Battista, M. T. 2010. “Engaging Students in Meaningful Mathematics Learning: Different Perspectives, Complementary Goals.” *Journal of Urban Mathematics Education* 3(2):34–46.

Baumert, J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, and Y.-M Tsai. 2010. “Teachers’ Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress.” *American Educational Research Journal* 47(1):133–180.

Boaler, J. 1997a. “Setting, Social Class and Survival of the Quickest.” *British Educational Research Journal* 23(5):575–595.

Boaler, J. 1997b. “When Even the Winners are Losers: Evaluating the Experiences of Top Set Students.” *Journal of Curriculum Studies* 29(2):165–182.

Boaler, J., D. Wiliam, and M. Brown. 2000. “Students’ Experiences of Ability Grouping: Disaffection, Polarization and the Construction of Failure.” *British Educational Research Journal* 26:631-648.

Burris, C. C., J. P. Heubert, and H. M. Levin. 2006. “Accelerating Mathematics Achievement Using Heterogenous Grouping.” *American Educational Research Journal* 43:105-136.

Carbonaro, W. 2005. “Tracking, Students' Effort, and Academic Achievement.” *Sociology of Education* 78(1):27-49.

Cheema, J. R. 2014. “A Review of Missing Data Handling Methods in Education Research.” *Review of Educational Research* 84(4): 487–508.

Chen, C.-H., M. D. Crockett, T. Namikawa, J. Zilimu, and S. H. Lee. 2012, “Eighth Grade

Mathematics Teachers' Formative Assessment Practices in SES-different Classrooms: A Taiwan Study.*" International Journal of Science and Mathematics Education* 10:553-579.

Cheng, R. W.-Y., D. M. McInerney , and M. M. C. Mok. 2014. "Does Big-fish–little-pond Effect Always Exist? Investigation of Goal Orientations as Moderators in the Hong Kong Context." *Educational Psychology* 34(5):561-580.

Chirinda, B., and P. Barmby. 2018. "South African Grade 9 Mathematics Teachers' Views on the Teaching of Problem Solving." *African Journal of Research in Mathematics, Science and Technology Education* 22(1):114-124.

Chmielewski, A. K. 2014. "An International Comparison of Achievement Inequality in Within- and Between-school Tracking Systems." *American Journal of Education* 120(3):293-324.

Chmielewski, A. K., H. Dumont, and U. Trautwein. 2013. "Tracking Effects Depend on Tracking Type: An International Comparison of Students' Mathematics Self-concept." *American Educational Research Journal* 50(5):925–957.

Conner, A. M., L. M. Singletary, R. C. Smith, P. A. Wagner, and R. T. Francisco. 2014. "Teacher Support for Collective Argumentation: A framework for Examining How Teachers Support Students' Engagement in Mathematics Activities." *Educational Studies in Mathematics* 86:401-429.

Creemers, B. P. M., and L. Kyriakides. 2008. *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London & New York: Routledge.

Day C., P. Sammons, A. Kington, E. Regan, J. Ko , E. Brown, … D. Robertson. 2008. *Effective*

Classroom Practice (ECP): A Mixed-method Study of Influences and Outcomes (Report submitted to the Economic and Social Research Council RES-000-23-1564). Nottingham: School of Education, University of Nottingham.

Dedrick, R., J. Ferron, M. Hess, K. Hogarty, J. Kromrey, and T. Lang et al. 2009. “Multilevel Modelling: A Review of Methodological Issues and Applications.” *Review of Educational Research* 79(1):69-102.

Domina, T., P. Hanselman , N. Hwang, and A. McEachin. 2016. “Detracking and Tracking Up: Mathematics Course Placements in California Middle Schools, 2003-2013.” *American Educational Research Journal* 53(4):1229-1266.

Dunne, M., J. Humphreys, J. Sebba, A. Dyson, F. Gallannaugh, and D. Muijs. 2007. *Effective Teaching and Learning for Pupils in Low Attaining Groups*. London: DCSF.

Ells, A., Z. Ozgur, and L. Reiten. 2019. “Teacher Moves for Supporting Student Reasoning.” *Mathematics Educational Research Journal* 31:107-132.

Else-Quest, N. M., J. S. Hyde, and M. C. Linn. 2010. “Cross-national Patterns of Gender Differences in Mathematics: A Meta-analysis.” *Psychological Bulletin* 136(1):103-127.

Forgasz, H. 2010. “Streaming for Mathematics in Years 7-10 in Victoria: An Issue of Equity?” *Mathematics Education Research Journal* 22(1):57-90.

Francis, B., L. Archer, J. Hodgen, D. Pepper, B. Taylor, and M.-C.Travers. 2017. “Exploring the Relative Lack of Impact of Research on ‘Ability Grouping’ in England: A Discourse Analytic Account.” *Cambridge Journal of Education* 47(1):1–17.

Gamoran, A. 2010. “Tracking and Inequality: New Directions for Research and Practice.” In *The Routledge International Handbook of the Sociology of Education*, edited by M. W. Apple, S. J. Ball, and L. A. Gandin, 213-228. London: Routledge.

Gentry, M. 2016. “Commentary on “Does Sorting Students Improve Scores? An Analysis of Class Composition”. *Journal of Advanced Academics* 27(2):124–130.

Hallam, S., and J. Ireson. 2005. “Secondary School Teachers’ Pedagogic Practices When Teaching Mixed and Structured Ability Classes.” *Research Papers in Education* 20(1):3–24.

Hattie, J., and H. Timperley. 2007. “The Power of Feedback.” *Review of Educational Research* 77(1):81-112.

Huang, H.-Y., and P.-J. Hsieh. 2019. “Survey on Implementation and Effectiveness of Ability Grouping in Eighth-grade Math Classes in Taiwan.” *International Journal of Intelligent Technologies and Applied Statistics* 12(4):417-428.

Johnston, O., and H. Wildy. 2016. “The Effects of Streaming in the Secondary School on Learning Outcomes for Australian Students : A Review of the International Literature.” *Australian Journal of Education* 60(1):42–59.

Juta, A., and C. Van Wyk. 2020. “Classroom Management as A Response to Challenges in Mathematics Education: Experiences from A Province in South Africa.” *African Journal of Research in Mathematics, Science and Technology Education* 24(1):21-30.

Klieme, E., C. Pauli, and K. Reusser. 2009. “The Power of Video Studies in Investigating Teaching and Learning in the Classroom. “ In *The Pythagoras Study*, edited by T. Janik and T. Seidel, 137–160. Münster: Waxmann.

Kunter, M., and J. Baumert. 2006. “Who Is the Expert? Construct and Criteria Validity of Student and Teacher Ratings of Instruction.” *Learning Environment Research* 9:231-251.

Larsson, M., and A. Ryve. 2012. “Balancing on the Edge of Competency-oriented Versus

Procedure-oriented Practices: Orchestrating Whole-class Discussions of Complex Mathematical Problems." *Mathematics Educational Research Journal* 24:447-465.

Lavie, I., A. Steiner, and A. Sfard. 2019. "Routines We Live By: From Ritual to Exploration." *Educational Studies in Mathematics* 101:153-176.

Lee, C. W., T. A. Walkowiak, and J. L. Nietfeld. 2017. "Characterization of Mathematics Instructional Practices for Prospective Elementary Teachers with Varying Levels of Self-efficacy in Classroom Management and Mathematics Teaching." *Mathematics Educational Research Journal* 29:45-72.

Liou, P.-Y., and Y.-C. Hung. 2015. "Statistical Techniques Utilized in Analyzing PISA and TIMSS Data in Science Education from 1996 to 2013: A Methodological Review." *International Journal of Science and Mathematics Education* 13:1449-1468.

Marais, C., G. Van Der Westhuizen, and H. Tillema. 2013. "Teacher Knowledge of Learners' Help-seeking in Mathematics Problem Solving." *African Journal of Research in Mathematics, Science and Technology Education* 17:1-2.

Markow, D., and M. Cooper. 2008. *The MetLife Survey of the American Teacher: Past, Present and Future – A Survey of Teachers, Principals and Students*. Retrieved November 9, 2017 from http://files.eric.ed.gov/fulltext/ED504457.pdf

Mazenod, A., B. Francis, L. Archer, J. Hodgen, B. Taylor, A. Tereshchenko, and D. Pepper. 2019. "Nurturing Learning or Encouraging Dependency? Teacher Constructions of Students in Lower Attainment Groups in English Secondary Schools." *Cambridge Journal of Education* 49(1):53-68.

McCrory, R., R. Floden, J. Ferrini-Mundy, M. D. Reckase, and S. L. Senk. 2012. "Knowledge of Algebra for Teaching: A Framework of Knowledge and Practices." *Journal for Research in Mathematics Education* 43(5):584–615.

Mijs, J. J. B. 2016. "Stratified Failure: Educational Stratification and Students' Attributions of Their Mathematics Performance in 24 Countries." *Sociology of Education* 89(2):137-153.

Nicol, C. 2002. "Where's the Math? Prospective Teachers Visit the Workplace." *Educational Studies in Mathematics* 50:289-309.

OECD. 2013. *PISA 2012 Results in Focus: What 15-year-olds Know and What They Can Do With What They Know*. Retrieved November 9, 2017 from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

OECD. 2014. *PISA 2012 Technical Report*. Retrieved November 9, 2017 from https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

Ofsted. 1993. *Access and Achievement in Urban Education.* London: Office for Standards in Education.

Osborne, J., S. Simon, and S. Collins. 2003. "Attitudes Towards Science: A Review of the Literature and Its Implications." *International Journal of Science Education* 25(9):1049-1079.

Paul, F. G. 2005. "Grouping within Algebra I: A Structural Sieve with Powerful Effects for Low-income, Minority, and Immigrant Students." *Educational Policy* 19:262-282.

Praetorius, A.-K., C. Pauli, K. Reusser, K. Rakoczy, and E. Klieme. 2014. "One lesson is all you need? Stability of instructional quality across lessons." *Learning and Instruction* 31:2-12.

Prediger, S., C. Fischer, C. Selter, and C. Schober. 2019. Combining Material- and Community-based Implementation Strategies for Scaling Up: The Case of Supporting Low-achieving Middle School Students. *Educational Studies in Mathematics* 102:361-378.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods (2nd ed.)*. Thousand Oaks: Sage.

Raudenbush, S., A. Bryk, Y. F. Cheong, R. Congdon, and M. Du Tolt. 2011. *HLM7: Hierarchical Linear and Nonlinear Modeling.* Lincolnwood: Scientific Software International.

Reynolds, D., P. Sammons, B. de Fraine, J. Van Damme, T. Townsend, C. Teddlie, and S. Stringfield. 2014. "Educational Effectiveness Research (EER): A State-of-the-art Review." *School Effectiveness and School Improvement* 25(2):197-230.

Robert, P. 2010. "Social Origin, School Choice, and Student Performance." *Educational Research and Evaluation* 16(2):107-129.

Schlesinger, L., and A. Jentsch. 2016. "Theoretical and Methodological Challenges in Measuring Instructional Quality in Mathematics Education Using Classroom Observations." *ZDM Mathematics Education* 48:29–40.

Schofield, J. W. 2010. "International Evidence on Ability Grouping with Curriculum Differentiation and the Achievement Gap in Secondary Schools." *Teachers College Record* 112(5):1490-1526.

Steenbergen-Hu, S., M. C. Makel, and P. Olszewski-Kubilius. 2016. "What One Hundred Years of Research Says about the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-order Meta-analyses." *Review of Educational Research* 86(4):849–899.

Tan, C. Y. 2020. *Family Cultural Capital and Student Achievement: Theoretical Insights from PISA*. Singapore: Springer.

Tan, C. Y., and D. Liu. 2017. "What is the Influence of Cultural Capital on Student Reading

Achievement in Confucian as Compared to Non-Confucian Heritage Societies?" *Compare* 48(6):896-914.

Taylor, B., B. Francis, N. Craig, L. Archer, J. Hodgen, A. Mazenod, A. Tereshchenko, and D. Pepper. 2019. "Why is it Difficult for Schools to Establish Equitable Practices in Allocating Students to Attainment 'Sets'? " *British Journal of Educational Studies* 67(1):5-24.

Tieso, C. L. 2003. "Ability Grouping is not Just Tracking Aymore." *Roeper Review* 26:29–36.

Touro, J., E. Navarro-Ascencio, L. Lizasoain, E. Loez-Gonzalez, and M. J. G.-S. Pedro. 2019. "How Teachers' Practices and Students' Attitudes Toward Technology Affect Mathematics Achievement: Results and Insights from PISA 2012." *Research Papers in Education* 34(3):263-275.

Wilkinson, S. D., and D. Penney. 2014. "The Effects of Setting on Classroom Teaching and Student Learning in Mainstream Mathematics, English and Science Lessons: A Critical Review of the Literature in England." *Educational Review* 66(4):411–427.

Xie, Y., M. Fang, and L. Shauman. 2015. "STEM Education." *Annual Review of Sociology* 41:331–357.

Yi, H. S., and Y. Lee. 2017. "A Latent Profile Analysis and Structural Equation Modeling of the Instructional Quality of Mathematics Classrooms Based on the PISA 2012 Results of Korea and Singapore." *Asia Pacific Education Review* 18:23-39.

Zevenbergen, R. 2003. "Grouping by Ability: A Self-fulfilling Prophecy." *Australian Mathematics Teacher* 59(4):2-7.

Zevenbergen, R. 2005. "The Construction of a Mathematical Habitus: Implications of Ability Grouping in the Middle Years". *Journal of Curriculum Studies* 37(5):607-619.

Table 1. Fixed effects estimates (top) and variance-covariance estimates (bottom) in three-level HLM.

	Model 1	Model 2	Model 3	Model 4	Model 5
		Fixed parameters			
Intercept	477.80***(5.98)	477.63***(4.67)	477.05***(4.69)	477.53***(4.69)	477.52***(4.66)
Student level					
Male		12.37***(1.79)	12.37***(1.80)	13.23***(1.79)	13.23***(1.79)
ESCS		19.39***(1.25)	19.38***(1.25)	19.20***(1.22)	19.20***(1.22)
COGACT				3.19***(0.64)	3.19***(0.64)
CLSMAN				3.99***(0.74)	3.99***(0.74)
TCHBEHFA				-8.30***(0.82)	-8.30***(0.82)
TEACHSUP				3.28**(1.02)	3.28**(1.02)
School level					
SomeClasses			-14.01***(3.67)	-13.74***(3.61)	-13.74***(3.53)
AllClasses			-22.03***(4.29)	-21.89***(4.22)	-22.02***(4.13)
Country level					
CHC		47.61***(8.28)	46.72***(9.73)	44.46***(10.09)	44.74***(9.97)
Cross-level interactions					
SomeClasses*COGACT					-5.12(3.58)
SomeClasses*CLSMAN					4.83(3.21)
SomeClasses*TCHBEHFA					5.13(3.05)
SomeClasses*TEACHSUP					-5.10(4.08)
AllClasses*COGACT					-7.00(4.41)
AllClasses*CLSMAN					1.47(2.76)
AllClasses*TCHBEHFA					3.35(3.14)
AllClasses*TEACHSUP					2.00(4.77)
		Random parameters			
Level 1 intercept	5,427.74	5,180.92	5,180.86	5,128.07	5,128.04
Level 2 intercept	3,827.05***	3,078.93***	3,011.84***	2,926.11***	2,911.67***
Level 3 intercept	1,047.37***	627.48***	659.81***	653.99***	651.03***
% Level 1 variance	52.69	58.30	58.52	58.89	59.01
% Level 2 variance	37.15	34.64	34.02	33.60	33.50
% Level 3 variance	10.17	7.06	7.45	7.51	7.49
% variance reduction as compared to Model 1					
Level 1		4.55	4.55	5.52	5.52
Level 2		19.55	21.30	23.54	23.92
Level 3		40.09	37.00	37.56	37.84

Note: Robust standard errors in parentheses.

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