Ability Grouping and Reading Achievement:
Mechanisms and effects on students in Hong Kong and 34 countries

Ming Ming Chiu, Bonnie Wing-Yin Chow, and Lawrence Khoo

Abstract of research (**limited to ½ A-4 page or 200 words**, and comprehensible to a non-specialist):

Hong Kong, like many countries, group their students by ability (tracking), hoping that tracked students learn more. We test whether tracked students learn more than untracked students in the area of reading. We also examine the mechanisms through which tracking might affect student learning: (1) modified teaching, (2) student interactions, (3) unequal allocation of resources, (4) achievement variance, and (5) self-concept. Then, we examine how tracking differs at the country, classroom, and student levels. In particular, we analyze how the link between tracking and learning differs across various types of students (smart vs. slow; rich vs. poor).

We use data from five sources. IEA-PIRLS gave reading tests to 146,490 primary school 4th graders from Hong Kong and 34 countries. Furthermore, IEA-PIRLS collected questionnaire responses from students, parents, teachers, and principals. We combine this data with four additional data sets: economic data from OECD (2000), and cultural values data from Bond et al (2004), Hofstede (2003) and Inglehart et al. (2004).

Our findings takes us a step closer to building a model of tracking and learning, which in turn can inform the Hong Kong government's and schools' policies on school banding, allocation of students to schools, and allocation of resources (e.g., money and teachers).

/.... PART II

PART II DETAILS OF THE RESEARCH PROPOSAL
[To be completed by the applicant(s)]

RESEARCH DETAILS

1(a). The project objectives and long-term impact **(maximum 1½ A-4 page)**:

Objectives

This study investigates whether primary school students grouped by past achievement (tracking) learn more than untracked students in the area of reading. We ask three major questions:

1. Does grouping students by ability (tracking) within or across schools affect overall student learning?

   Specifically, do tracked students learn more than non-tracked students in the area of reading?
2. What are the mechanisms by which tracking affects student learning?
3. Are these effects context-dependent? Do they differ across countries, schools or students?

   Smarter vs. weaker students?

   Richer vs. poorer students?

Theoretical significance

1. Help build a theory of how different types of tracking affect different students' learning.
2. Examine the moderating influences of country, school, classroom, and family contexts.

   Examining other countries shows whether links are universal or specific to Hong Kong
3. Show how tracking effects might structurally differ across students (Smarter vs. weaker? Richer vs. poorer?).

Practical significance

Findings of this study will inform Hong Kong government's and schools' policies on

1. Allocation of students
   a. Across schools, including size of school bands
   b. Within schools
   c. Impact of allocation criteria
2. Allocation of resources
   a. Government funding for different types of schools
   b. Budgeting of resources within each school

Improving allocation of students or resources costs little.
Yet, improved allocation can yield substantial learning benefits.

Methodological significance

To test the above policy questions, we address several major difficulties as follows:

Methodological Difficulties	Solutions
Efficient and precise test design?	Balanced incomplete block (BIB) tests 3-parameter logistic (3PL) item response model
Representative sampling?	Stratified sampling across schools and students
Questionnaire measurement error?	Warm indices
Missing questionnaire response data?	Markov Chain Monte Carlo multiple imputation
Nested data? Students in classes in schools in countries	Multi-level analyses (hierarchical linear modeling, HLM)

Our integration of these statistical methods can serve as primers for future analyses of complex data.

Early release of results

Time available after project starts	Information and Dissemination Format & Methods	Potential Users/ Audience
1 year	Preliminary results of tracking and achievement in Hong Kong Press release & press conference Website report	Government officials, Principals, Teachers, Parents, General public
1 year 6 months	Preliminary results of tracking and achievement in 35 countries Press release & press conference Website report	Government officials, Principals, Teachers, Parents, General public

/.... 1(b). Policy

1(b). Policy Implications of the research proposal **(maximum 1 A-4 page)**:

This study on ability grouping of students (tracking) and resource allocation can help improve student learning by informing the Hong Kong government and schools' policies on allocations of students and resources (e.g., teachers). See table A. As these issues regard allocation, implementation of these policy implications do not require extra spending, and might yield substantial improvement in Hong Kong students' learning.

Table A. The policy implications of findings from this study's research questions.

Study questions	**Policy Implications**
1) What types of tracking improve student learning? (if any) Which students, if any, benefit from tracking?	Hypothetical Examples: a. If tracking 3 bands of schools improve students learning, then HK-EMB should maintain the current banding system. b. If tracking does not improve student learning, then HK-EMB should eliminate the current banding system c. If grouping students by past achievement within classes improves overall scores, then HK-EMB should encourage schools to group students by past achievement within a class.
2) Do some types of schools use some types of resources(e.g. teachers) substantially more effectively than others?	Hypothetical Examples: a. If students in government schools benefit more from school resources than students in other schools, then HK-EMB should allocate more school resources to government schools. b. If students from different types of schools show similar benefits from school resources, then HK-EMB should maintain the current system of equal funding for each student.
3) Do some types of students use resources substantially more effectively than others?	Hypothetical Examples: a. If students showing low past achievement learn more from teachers with more years of teaching experience, then HK-EMB should encourage schools (e.g. with incentives) to assign more experienced teachers to students with lower past achievement. b. If students showing high past achievement learn more from teachers with more years of teaching experience, then HK-EMB should encourage schools (e.g. with incentives) to assign more experienced teachers to students with higher past achievement.

Hong Kong government policy influences student achievement by allocating both students and school resources. The Hong Kong government's Education and Manpower Bureau (HK-EMB) decides the school that many students attend, 47% of primary school students and for 81% of secondary students in 2005 (for enrolling in government and aided schools only, HK-EMB, 2005a, 2006). In 2004-2005, HK-EMB spent $54 billion (23% of her total expenditure) on education, $10 billion on primary schooling and $16 billion on secondary schooling (HK-EMB, 2006). Lastly, HK-EMB hires and allocates all primary and secondary school teachers in government schools (HK-EMB, 2005b).

Which type of tracking, if any, improves student learning? For which students? There are three possible effects of the current tracking system. First, the top, middle, and bottom third of students might benefit most from similar students, such that tracking yields higher overall academic achievement. Then, the current system is optimal, and no changes should be

made. Second, if tracking does improve learning, but the current school band sizes are not optimal, then the school band sizes should be changed accordingly. Third, if tracking across schools does not improve overall student learning, then there should be no separate school bands. In this case, HK-EMB should not allocate school places based on students' past academic performance, and they should use other methods, such as random allocation.

Likewise, if ability grouping within a class substantially improves overall student learning, then the Hong Kong government can encourage schools to group their students by ability within each classroom. On the other hand, if ability grouping within a class does not improve overall student learning, then the Hong Kong government should discourage schools from using such practices.

In addition, which schools, if any, should get more students and more resources? If some types of schools (e.g., high vs. middle vs. low bands) use specific resources (e.g. highly qualified teachers) substantially more effectively than others to improve their student learning, then the government should consider allocating more students and resources to those schools. On the other hand, if the differences are not significant, then each school should receive the same amount of resources per student.

Likewise, should schools be required to give similar resources to each student? If some types of students (rich vs. poor; smart vs. slow) learn substantially more per unit of a given resource (e.g., education materials) than other students, then schools should consider allocating more of that specific resource to these types of students. On the other hand, if the differences are not significant, then each student should receive the same amount.

All policy recommendations entail the re-allocation of existing students and school resources. Thus, no extra money is needed for new programs. Hence, low-cost re-allocation of current resources might yield substantial improvement in student learning.

[All references appear at the end of the research plan and methodology section.]

/.... 2. Background

2. Background of research **(maximum 2½ A-4 pages, including references)**:

Tracking (ability grouping or streaming) creates more homogeneous groups of students by assigning students to schools, classes, or groups within a class on the basis of their past achievement. The effects of tracking on learning differed across studies (*positive*: e.g., Kulik & Kulik, 1982, 1987; *negative*, e.g., Hanushek & Wößmann, 2005; *non-significant*: Opdenakker & Van Damme, 2001; Slavin, 1990). These conflicting findings suggest that tracking might operate through multiple mechanisms and differ across countries, schools or students. Tracking might affect student learning through several mechanisms such as (1) modified teaching, (2) student interactions, (3) unequal allocation of resources, (4) achievement variance, or (5) self-concept. Furthermore, these mechanisms might differ across schools or across countries. Their effects might also differ across various types of students (rich vs. poor; smart vs. weak). By testing how tracking operates across countries, schools, and students, this study can help explain different tracking effects and inform government and school policies to improve student learning.

Teaching tracked students

Adapt instruction

In response to tracking, teachers might adapt their instruction to groups of students and/or stereotype their students. By creating groups of students with similar past achievement, tracking helps educators customize the curriculum, lessons, teaching materials, and teaching pace to the needs of each group of students, which might improve students'

learning (Hallinan & Kubitschek, 1999; Liu, Wang & Parkins, 2005; Slavin, 1987a). Consider a student on a high track, Heidi, and a student on a low track, Lola. During tracking, Heidi enjoys more difficult lessons that stimulate and challenge her. However, these difficult lessons might frustrate Lola. Instead, Lola attends easier lessons with attainable goals. Hence, tracking might raise overall student learning by adapting instruction to the needs of each group of students (see table 1, middle column, hypothesis 1-a).

Table 1. Possible mechanisms through which tracking might affect achievement.

Tracking		**Hypothesized Tracking Mechanisms**		**Reading**
across classes		1-a) Adapt instruction (+)		**Achievement**
		1-b) Stereotyping (–)		
within classes				
		2-a) Diminishing marginal returns (–)		
		2-b) Smart use (+)		
		3-a) Homophily (+)		
		3-b) Less Diversity (–)		
		3-c) Less student help (–)		
		4) Mass Market (–)		
		5-a) Labeling (smart: +; weak: –)		
		5-b) Big fish, little pond (smart: – ; weak: +)		

Stereotyping

Tracking affects teacher expectations of students and might yield stereotyping of students in high tracks as smart (Heidi) and those in low tracks as not smart (Lola). As a result, Heidi's high track teachers tend to value academic teaching goals more, view their students as more teachable, and have higher academic expectations of them (Pallas, Entwisle, Alexander, & Stulka, 1994; Van Houtte, 2004). These higher teacher expectations tend to boost Heidi's confidence and help her attain higher goals (Hallinan & Kubitschek, 1999). In contrast, teachers with low expectations of Lola might give her insufficiently challenging classwork, which can result in both less learning and reduced motivation to learn. Regardless of the track, teachers might tend to stereotype students based on their track and be less sensitive to the needs of individual students. For example, teachers might tend to overlook Heidi's weaknesses or underestimate Lola's strengths during tracking. As a result, tracking might tend to lower overall student achievement (see table 1, hypothesis 1-b).

Unequal allocation of resources

Previous research findings have suggested that tracking favors high-achieving students (e.g., Gamoran, 1987; Hallinan & Kubitschek, 1999; Hoffer,1992; Kulik & Kulik, 1982; Opdenakker & Van Damme, 2001) while harming the low-achieving ones (Hallinan & Kubitschek, 1999; Hoffer,1992; Galloway & Schwartz, 1994; Oakes, 1982, 1985; Vanfossen, Jones, & Spade, 1987). These differential effects might stem from the unequal allocation of resources that accompanies tracking. At minimum, tracking creates unequal allocation of smart classmates. Often unequal allocation of other educational resources also occurs during tracking. This greater inequality might result in less efficient use of resources and lower academic achievement. Or, it might result in better use of resources by smarter students and higher academic achievement.

More resources to smarter students

Viewing students as resources, tracking gives high track students like Heidi more resources in the form of her smarter groupmates or classmates. In contrast, Lola's groupmates or classmates are not as smart. For example, Lola has fewer successful role models, especially during tracking across classes (Eder, 1981; Eder & Felmlee, 1984). As smarter students are also often better disciplined, high track students often enjoy better school climate and more disciplined classes or groups (Berends, 1995). In low track classes or groups on the other hand, students like Lola faces more student misbehavior, which can reduce her sense of psychological safety and comfort in the classroom (Eder, 1981; Eder & Felmlee, 1984).

Better teachers are also attracted to the smarter students in higher tracks. Not surprisingly, teachers in higher tracks have higher academic qualifications, more years of schooling and more teaching experience (Oakes, 1985). These positive teachers' personal qualities are all linked to more efficient instructional practices employed in higher tracks. These teachers tend to have better prepared lessons and promote student learning with more zeal (Oakes 1985; Rosenbaum 1976). In higher tracks, teachers use more complex teaching materials (Oakes, 1990; Page, 1987), include more academic courses (Gamoran, 1987; Vanfossen, Jones, & Spady, 1987), teach at a faster pace (Oakes, 1990), and lead more stimulating discussions (Grossen, 1996). Thus, high track students like Heidi learn more. In contrast, students like Lola receive fewer explanations and directions regarding teacher expectations and goals for them, which reduce her learning opportunities (Evertson 1982; Goodlad 1984; Oakes 1985). These distinct instructional practices might explain the differences in academic performance across high and low track groups (Gamoran, 1989; Grossen, 1996; Van Houtte, 2004).

More resources to richer students

Tracking across schools or classes creates elite schools/classes that rich students seek to attend. Hence, parents of high SES families tend to advocate tracking and place their children in high track schools or classes (Opdenakker & Van Damme, 2001), though Marks (2006) found no such link. Richer students often attend high track schools that have more resources, better physical conditions and better education materials (e.g., Berner, 1993; Chiu & Khoo, 2005; Fuller & Clarke, 1994). High-track classes also have proportionately more high SES students (Gamoran, 1987; Opdenakker & Van Damme, 2001). Students of families with higher *SES* have better educational resources (Entwisle & Alexander, 1995), larger social networks with more highly skilled or educated people (Horvat, Weininger, & Lareau, 2003), and more human capital which helps them to acquire cognitive and social skills, and social and cultural norms more effectively (e.g., Ochs, Taylor, Rudolph, & Smith, 1992). Thus, high track students like Heidi often have classmates from higher SES families. These classmates tend to understand and appreciate the value of schooling more, thereby contributing to a class or school culture of higher academic achievement and better discipline (Chiu & McBride-Chang, 2006; Davalos, Chavez, & Guardiola, 2005; Willms, 1999). As a result, high track students like Heidi often enjoy more learning resources from their richer classmates or groupmates, while Lola gets fewer learning resources (Chiu & Khoo, 2005). These inequalities during tracking might (a) reduce overall academic performance due to diminishing marginal returns, or (b) increase overall academic performance owing to smart use.

Diminishing marginal returns

Like many resources, educational resources yield *diminishing marginal returns* (as a thirsty person benefits more from the first glass of water than from the last glass of water, Chiu & Khoo, 2005; Dar & Resh, 1986). For example, an extra $100 in books likely

improves a poor student's reading score more than a rich student's reading score. As poorer students receive more resources in more equal situations and benefit from them more than do richer students, schools or countries with more equal distribution of resources have higher overall achievement (Chiu & Khoo, 2005). As tracking tend to increase inequality of educational resources among students, tracking might lower overall achievement due to diminishing marginal returns (see table 1, hypothesis 2-a).

Smart use

Or, high-achieving students might use resources (e.g., a book, smart schoolmate) more effectively than other students (Dillon & Watson, 1996). A high-achieving student like Heidi might read a storybook and understand more compared to Lola. As smarter students might benefit more from extra educational resources, tracking might raise overall student achievement (*smart use*, see table 1, hypothesis 2-b).

Interactions among tracked students

Unequal resource distribution might also hinder students from sharing resources and ideas with one another (Goldsmith, 2004). Hence, tracking might affect students' interactions through homophily, diversity, and student help.

Homophily

People prefer to befriend and interact with others of the same ability, socio-economic status [SES], gender, and age (*homophily*, McPherson, Smith-Lovin, & Cook, 2001). Thus, students like Heidi might befriend more people like herself in homogeneous groups, classes or schools and thereby share more resources, compared to students in heterogeneous ones. Likewise, teachers tend to prefer interacting with students like themselves, and thus tend to favor high track students like Heidi and middle-class or richer students (cultural gate-keeping, Roscigno & Ainsworth-Darnell, 1999). According to the homophily hypothesis, clustering similar people together tends to raise each student's achievement, while allocating higher quality teachers to low track students like Lola tends to lower overall student achievement (see table 1, hypothesis 3-a). However, student resentment at the inequality within a school can reduce cooperation and hinder student learning (Goldsmith, 2004).

Less diversity

On the other hand, people learn more from one another's differences in experiences than from one another's similarities (cf. workplace diversity effects, Howard & Brakefield, 2001). As smarter and weaker students (or richer and poorer students) tend to have different experiences, Heidi and Lola might learn more from each other than from students similar to themselves. As tracking reduces *diversity* within a class or a group, it might lower overall student achievement (see table 1, hypothesis 3-b)

Less student help

In a mixed class, smart students like Heidi have more opportunities to explain ideas to low-achieving students like Lola, benefiting both sides (Webb & Palincsar, 1996). Lola gets extra help. Meanwhile, Heidi elaborates her understanding by explaining to Lola (Webb, 1991). However, separating high and low ability students reduces these kinds of *helping* opportunities. As tracking decreases helping and learning opportunities of both high and low ability students, it might lower overall student achievement (see table 1, hypothesis 3-c)

Greater variance and mass market

Tracking reduces the variance within a group, class or school. However, the clustering of high-achieving students and of low-achieving students might create proportionately more

students at each extreme –extremely high achievers and extremely low achievers. As these students often have special needs, meeting these needs are often costly, requiring special skills or materials. As media and sellers target the largest common population segment and cultural background knowledge (*mass market*), most resources are geared toward the large majority of middle-achieving students, rather than the special needs of the highest- or lowest-achieving students (e.g., Crowther, Dyson, & Millward, 2001). The high costs for helping students with special needs take away limited educational resources from other students. Or, these special needs might not be met. In either case, overall achievement would be lower. Hence, tracking might reduce overall achievement by creating students with more costly, special needs according to the mass market hypothesis (see table 1, hypothesis 4).

Academic self-concept during tracking

Labeling

Tracking might also influence students' academic achievement indirectly through their academic self-concept: labeling or social comparison. A low track label stigmatizes students like Lola as low achievers (Liu, Wang & Parkins, 2005). This stigma tends to lower Lola's academic self-concept and her motivation to learn. On the other hand, Heidi enjoys the reflected glory of a high track label: higher self-concept, higher motivation to learn, and positive attitudes toward schools (Berends, 1995; Kulik & Kulik, 1982). This labeling effect might vary across between-class and within-class tracking due to the reference groups available to students. Labeling effects are often stronger in within-class tracking rather than across-class tracking (Reuman, 1989).

Big fish, little pond

Social comparison might also affect students' academic achievement during across-class tracking (but not during within-class tracking). When surrounded by lower-achieving classmates, low track students like Lola often have higher academic self-concept, which in turn raises their academic achievement (big fish, little pond effect, Liu, Wang & Parkins, 2005; Marsh & Hau, 2004). In contrast, Heidi has smarter classmates, which might lower her academic self-concept and achievement. These tracking effects are further complicated by the effects of time. Liu, Wang & Parkins (2005) found that low-achieving groups of students had a more negative academic self-concept immediately after tracking but more positive academic self-concept three years after tracking, compared to high-achieving groups.

Other variables linked to academic achievement

Liu, Wang and Parkins (2005) found that differences in girls' scores across high and low tracks accounted for most of the tracking effects. They speculate that girls are more socially sensitive to their classmates, compared to boys. However, other studies have not shown gender differences (e.g., Opdenakker & Van Damme, 2001).

We also statistically controlled several other variables that are linked to student achievement: immigrant status (Fuligni, 1997; Portes & MacLeod, 1996), language spoken at home (Collier, 1995), family SES (Bradley & Corwyn, 2002), schoolmate SES (Hanushek et al., 2003), school properties (Baker, et al, 2002), teacher properties (Chiu & Khoo, 2005), and reading enjoyment (Chiu & McBride-Chang, 2006).

As shown above, the effects of tracking might differ across classes, schools and countries due to curriculum differences, amount of resources diverted to high track schools or classes, student interactions, and type of tracking. Thus, we also examine the effects of tracking in each class, in each school and in each country.

[All references appear at the end of the research plan and methodology section.]

3. Research plan and methodology **(maximum 3 A-4 pages, including key references)**:

This study investigates whether tracking affects primary school student learning reading in Hong Kong and in 34 countries.

1. Does grouping students by ability (tracking) within or across schools affect overall student learning?
   Specifically, do tracked students learn more than non-tracked students?
2. What are the mechanisms by which tracking affect student learning?
3. Do these effects differ across countries, schools or students?
   Smarter vs. slower students?
   Richer vs. poorer students?

To address the above questions across many countries and schools, we must address the following major difficulties: test design, representative sampling, questionnaire measurement error, missing data, and clustered data from multiple levels. We will do so by using (a) balanced incomplete block tests and a graded Rasch model, (b) stratified sampling across schools and students, (c) Warm estimated indices, (d) Markov Chain Monte Carlo imputation, and (e) multi-level analyses.

The International Association for the Evaluation of Educational Achievement's Progress in International Reading Literacy Study (PIRLS) assessed the reading literacy of 146,490 primary school 4th graders from 35 countries. The data also include questionnaire responses from students, parents, teachers, and principals. We will also use four additional data sets: economic data from OECD (2000), and cultural values data from Bond et al (2004), Hofstede (2003) and Inglehart et al. (2004).

International experts from participating countries defined levels of reading proficiency, built a framework for assessing it, created test items, forward- and backward-translated them, and piloted them to ensure validity and reliability (for details including reliability, validity checks, sample assessment items, and questionnaire items, see IEA, 2003).

Variables

The following variables will be obtained or computed from the PIRLS data unless otherwise indicated. Continuous variables will be standardized to means of zero and standard deviations of one (for details on these variables, see IEA, 2003).

Reading achievement. IEA (2003) estimated each student's reading achievement with a three-parameter logistic (3PL) item response model of student responses to multiple choice questions (Lord, 1980) and a 2PL generalized partial credit model for student responses to questions that allowed for partial credit (polytomous items; Muraki, 1992).

Past achievement. Student scores on 7 reading skills before entering primary school. We also compute the mean past achievement of each classroom. By modeling this measure, we can estimate the size of classmate effects to test the homophily, diversity, and student help hypotheses. To test the smart use hypothesis, we model interactions between past achievement and country, family, school, and schoolmate resource variables below. These measures also help compute the degree of tracking in a country.

Country income. We will use *real GDP per capita*, adjusted for inflation (data source: OECD, 2000) to represent country resources. To test the diminishing marginal returns hypothesis, we will also test log real GDP per capita as well as the logs of the family, school, and schoolmate resource variables below.

Cultural values. Using the three cultural value databases, we test the following cultural values: hierarchy/egalitarian, collectivist/individualist, masculine/feminine, uncertainty avoidance/ risk tolerance, short-term/long-term orientation, traditional/secular, survival/self-expression, dynamic externality, and social cynicism (Bond et al., 2004; Hofstede, 2003; Inglehart et al., 2004).

Family. SES indicators (*mothers' years of schooling, fathers' years of schooling, family income,* and *highest job status of parents*) may reflect one or many phenomena (Ostrove, Feldman, & Adler, 1999). Thus, we test inter-correlations among these variables. Based on the results, we will create an SES index or allow for separate effects in the regressions by entering each one separately and then together. Job status is measured using Ganzeboom, de Graaf, and Treiman's (1992) index. We will also compute the *class means* of the SES indicators to measure the SES of a student's classmates. Other family variables include first generation immigrant, second generation immigrant, foreign language spoken at home, family socio-economic status, and bilingual home environment.

Gender. We will test and control for the effect of *girl*, and *% of girls in a class*.

Measures of classmate distribution by past achievement and family SES. We have obtained income inequality data for each country, *GDP Gini* (data source: OECD, 2000). We will compute the prior reading skill variance within each country, within each class, and across classes. The ratio of the prior reading skill variance across classes over total prior reading skill variance within a country indicates the degree of ability grouping (clustering of smarter students away from other students). Likewise, we will also compute the total family SES variance within each country, within each class, across classes, and the clustering of rich students.

Tracking. We measure systematic ability grouping (tracking) in 4 ways: (a) grade level at which tracking formally begins in a country, (b) degree of clustering of students by past achievement within classes in a country, (c) tracking policy of the school, and (d) ability grouping within classrooms.

School. We test the following additional school characteristics: availability of school resources, school violence, school safety, and school climate.

Teacher. We will test the following classroom teacher properties: age, gender, years of education, teaching certification, academic preparation for teaching reading, teaching experience, and professional development. To test our stereotyping hypothesis, we model teacher expectations of students and responsiveness to weaker students.

Classroom activities. To test our adapt instruction hypothesis, we will model the following classroom characteristics: classroom reading time, classroom activities, and instruction in reading skills.

Student. Student variables include view of school, reading attitude, reading self-concept, and reading time. We compute the class means of these variables to capture classmate properties. To test the labeling and big fish, little pond hypotheses, we model the effects of class mean reading self-concept.

Analysis

We will model reading scores with the following multi-level specifications, using MLn software (Rasbash & Woodhouse, 1995). Ordinary least-squares regressions tend to underestimate the standard errors of regression coefficients. In contrast, multi-level models separate unexplained error into student (level 1), class (level 2), and country (level 3) components, thereby removing the correlation among error terms resulting from the nested data (students nested within classes within countries). We begin with a variance components model to test if the variances are significant at each level.

$$Y_{ijk} = \beta_{000} + e_{ijk} + f_{0jk} + g_{00k} \quad (1)$$

Y_{ijk} is the achievement score for student *i* in class *j* in country *k*. β_{000} is the grand mean intercept. The error terms (residuals) at the student-, class-, and country-levels are e_{ijk}, f_{0jk}, and g_{00k}, respectively.

We modeled students' achievement scores with sequential sets of variables (also known as hierarchical sets, Cohen, et al., 2003) to estimate the variance explained by each set. Past achievement is entered first as a control variable to predict improvement between past and present achievement, namely learning. By focusing on the smaller time period between past achievement and current achievement rather than a student's entire life, this approach reduces the problems of estimating the effects of family variables that might change over time (e.g. parents' jobs might change year to year; Hanushek, Cain, Markman & Rivkin, 2003).

Country variables might affect family variables (whereas a single family is unlikely to substantially affect country variables like GDP per capita or cultural values). Family variables, gender, and classmate family variables might affect choice of schools. All of these might affect student achievement. Hence**,** we entered the variables in this order: past achievement, country variables, family variables, classmates' family variables, school variables, class variables, and student variables.

We began by entering the past achievement variable: past reading skills (**T**).

$$Y_{ijk} = \beta_{000} + e_{ijk} + f_{0jk} + g_{00k} + \beta_{tjk}\mathbf{T}_{ijk} \quad (2)$$

Next, we enter *u* country level variables: log GDP per capita, total country variance of highest parent job status, cultural values of egalitarianism, individualism, masculinity, uncertainty avoidance, short-term orientation, traditional/secular, survival/self-expression, dynamic externality, social cynicism, start of tracking grade level, clustering of classmates by highest parent job status, and clustering of classmates by prior achievement (**U**).

$$Y_{ijk} = \beta_{000} + e_{ijk} + f_{0jk} + g_{00k} + \beta_{tjk}\mathbf{T}_{ijk} + \beta_{00u}\mathbf{U}_{00k} \quad (3)$$

We tested whether these variables were significant with a nested hypothesis test (χ^2 log likelihood, Cohen & Cohen, 1983). Non-significant variables were removed.

Next, we added *v* family variables, first generation immigrant, second generation immigrant, foreign language spoken at home, family socio-economic status, and bilingual home (**V)**.

$$Y_{ijk} = \beta_{000} + e_{ijk} + f_{0jk} + g_{00k} + \beta_{tjk}\mathbf{T}_{ijk} + \beta_{00u}\mathbf{U}_{00k} + \beta_{vjk}\mathbf{V}_{ijk} \quad (4)$$

As with **U**, we did nested hypothesis tests for **V**, removing non-significant variables. Next, we added *w* classroom/school variables (**W**): classmates' mean family SES, classmates' mean past achievement, availability of school resources, tracking within school, school violence, school safety, school climate; classroom teacher's age, gender, years of education, teaching certification, academic preparation for teaching reading, teaching experience, professional development, expectations of students, responsiveness to weaker students; classroom reading time, classroom activities, instruction in reading skills; same ability grouping within class, mixed ability grouping within class; classmates' views of school, attitude toward reading, reading self-concept, and reading time.

$$Y_{ijk} = \beta_{000} + e_{ijk} + f_{0jk} + g_{00k} + \beta_{tjk}\mathbf{T}_{ijk} + \beta_{00u}\mathbf{U}_{00k} + \beta_{vjk}\mathbf{V}_{ijk} + \beta_{0wk}\mathbf{W}_{0jk} \quad (5)$$

Next, we tested if the *v* student-level regression coefficients ($\beta_{vjk} = \beta_{v00} + f_{vjk} + g_{v0k}$) and the *w* school-level regression coefficients ($\beta_{0wk} = \beta_{0w0} + g_{0wk}$) differed significantly at the school-level ($f_{ujk} \neq 0$?) or the country-level ($g_{u0k} \neq 0$? or $g_{0wk} \neq 0$?). If so, we tested whether larger or smaller effects (slope) significantly correlated with higher reading achievement (intercept), and estimated the correlation of the slope-intercept relationship. We will estimate the correlation (r) between the size of the advantage and the average scores (Goldstein, 1995). We assume that the pairs of intercepts and slopes of *u* student-level variables have a bivariate

normal distribution with mean zero across the population of classes within a country, $E(f_{0jk}) = 0$. $Var(f_{0jk}) = t_{00k}$, $Var(f_{ujk}) = t_{uuk}$. The covariance of the class intercepts and the slopes is $Cov(f_{0jk}, f_{ujk}) = t_{0uk}$. Then, their correlation is

$$r(f_{0jk}, f_{ujk}) = t_{0uk} / (t_{00k} t_{uuk})^{0.5} \quad (6)$$

The computations are similar for countries and the *u* student-level variables and *w* class-level variables.

Lastly, we entered *x* student variables: view of school, reading attitude, reading self-concept, and reading time (**X**).

$$Y_{ijk} = \gamma_{000} + e_{ijk} + f_{0jk} + g_{00k} + \gamma_{tjk}\mathbf{T}_{ijk} + \gamma_{00u}\mathbf{U}_{00k} + \gamma_{vjk}\mathbf{V}_{ijk} + \gamma_{0wk}\mathbf{W}_{0jk} + \gamma_{xjk}\mathbf{X}_{ijk} \quad (9)$$

We will test the interactions of all significant variables. For example, interactions of variables with students' past achievement or family SES will test the differences between high vs. low ability students or rich vs. poor students.

To facilitate interpretation of the results, we will report how a ten percent increase in each continuous predictor above its mean affected reading scores (10% effect = b * SD * [10% / 34%]; 1 SD » 34%). In our base-10 system 10% is a common reference number. Note that scaling a 10% increase is not warranted as the percentage increase is not linearly related to the standard deviation.

We will use an alpha level of .05 for all statistical tests. As many significance tests on the same data can yield some false rejections of null hypotheses, Benjamini, Krieger, and Yekutieli's (in press) computer simulations showed that the two-stage linear step-up procedure addresses this issue better than 13 other methods. A multi-level version of the Sobel (1982) mediation test will be used (Krull & MacKinnon, 2001).

To test the mass market hypothesis, we will repeat the above analysis on the top and bottom 5%, 10%, and 25% of students by (a) past achievement in each country and (b) family SES in each country.

As readers might have interest in specific countries, we will do separate sets of two-level regressions (class and student levels) for each of the 41 countries with all variables except the country-level variables (Goldstein, 1995).

References

Baker, D. P., Goesling, B., & Letendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review, 46*, 291-312.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (in press). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*.

Berends, M. (1995). Educational stratification and students' social bonding to school. British *Journal of Sociology of Education, 16*(3), 327-352.

Berner, M. (1993). Building conditions, parental involvement, and student performance in the district of Columbia public school system. *Urban Education, 28*(1), 6-29.

Bond, M. H., Leung, K., Au, A., Tong, K.-K., De Carrasquel, S. R., Murakami, F., et al. (2004). Culture-level dimensions of social axioms and their correlates across 41 cultures. *Journal of Cross-Cultural Psychology, 35*(5), 548-570.

Chiu, M. M., & Khoo, L. (2005). Effects of resources, inequality, and privilege bias on achievement: Country, school, and student level analyses. *American Educational Research Journal, 42*, 575-603.

Chiu, M. M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 41 countries. *Scientific Studies of Reading.*

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral mathematics*. Hillsdale, NJ: Erlbaum.

Cohen, J., West, S. G., Aiken, L., & Cohen, P. (2003). *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum.

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal, 23*, 357-374

Davalos, D. B., Chavez, E. L., & Guardiola, R. J. (2005). Effects of perceived parental school support and family communication on delinquent behaviors in Latinos and white non-Latinos. *Cultural Diversity and Ethnic Minority Psychology, 11*(1), 57-68.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education, 54,* 151-161.

Eder, D., & Felmlee, D. (1984). The development of attention norms in ability groups. In P. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 189-208). New York: Academic.

Entwisle, D. R., & Alexander, K. L. (1995). A parent's economic shadow. *Journal of Marriage and the Family, 57*, 399-409.

Evertson, C. (1982). Differences in instructional activities in higher and lower-achieving junior high English and Math classes. *Elementary School Journal, 82,* 329-350.

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research, 64,* 119-157.

Galloway, D., & Schwartz, W. (1994). *Designing more effective grouping practices at the high school level.* Paper presented at the Annual Meeting of the Association for Supervision and Curriculum Development. Chicago, March 1994.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education,* 60, 145-55.

Gamoran, A. (1989). Measuring curriculum differentiation. *American Journal of Education,* 97, 129-143.

Ganzeboom, H. B. G., De Graaf, P., & Treiman, D. J. (with De Leeuw, J.). (1992). A standard international socio-economic index of occupational status. *Social Mathematics Research, 21*, 1-56.

Goldsmith, P. A. (2004). Schools' role in shaping race relations: Evidence on friendliness and conflict. *Social Problems, 51*, 587-612.

Goldstein, H. (1995). *Multilevel statistical models*. Sydney: Edward Arnold.

Grossen, B. (1996). *How shall we group to achieve excellence with equality.* Eugene, OR: University of Oregon.

Hallinan, M. & Kubitschek, W. (1999). Curriculum differentiation and high school achievement. *Social Psychology of Education, 3,* 41-62.

Hanushek, E. A., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differneces-in-differences evidence across countries. *The Economic Journal, 116,* C63-C76.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*, 527-544.

Heston, A., Summers, R., & Aten, B. (2002). *Penn world table version 6.1*. Philadelphia, PN: Center for International Comparisons at the University of Pennsylvania (CICUP).

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic performance across twenty-nine high- and low-income countries. *American Journal of Sociology, 88*, 1162-1194.

Hoffer, T. B. (1992). Middle school ability grouping and student school achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14,* 205-227.
Hofstede, G. (2003). *Culture's consequences*. Thousand Oaks, CA: Sage.
Hong Kong Education and Manpower Bureau (2005a). *Report on primary one admission 2005*. Retrieved November 8, 2006, from http://www.emb.gov.hk.
Hong Kong Education and Manpower Bureau (2005b). *Report on review of medium of instruction for secondary schools and secondary school places allocation*. Retrieved November 8, 2006, from http://www.emb.gov.hk.
Hong Kong Education and Manpower Bureau (2006). *Information leaflet on the Secondary School Places Allocation 2005/2007 cycle*. Retrieved November 8, 2006, from http://www.emb.gov.hk.
Hong Kong Government. Government expenditure on education. Retrieved October 27th, 2006 from www.emb.gov.hk.
Horvat, E. M., Weininger, E. B., & Lareau, A. (2003). From social ties to social capital. *American Educational Research Journal, 40*, 319-351.
Inglehart, R., Basanez, M., Diez-Medrano, J., Halman, L., & Luijkx, R. (2004). *Human beliefs and values*. Mexico City: Siglo XXI.
International Association for the Evaluation of Educational Achievement [IEA]. (2003). *PIRLS 2001 Technical Report*. Boston, MA: Boston College.
Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249-277.
Kulik, C. L. & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19,* 415-428.
Kulik, C. L. & Kulik, J. A. (1987). Effects of ability grouping on student achievement. *Equality and Excellence, 23,* 22-30.
Kulik, J. A. & Kulik, C. L. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36,* 73-77.
LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal, 40*(1), 43-89.
Liu, W. C., Wang, C. K. J., & Parkins, E. J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology, 75*(4), 567-586.
Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.
Marks, G. N. (2006). Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research, 48*(1), 21-40.
Marsch, H. W., & Hau, K. T. (2004). Explaining Paradoxical Relations Between Academic Self-Concepts and Achievements: Cross-Cultural Generalizability of the Internal/External Frame of Reference Predictions Across 26 Countries. *Journal of Educational Psychology, 96*(1), 56-67.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-44.
Muraki, E. (1992). A generalized partial credit model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*, 159-76.
Oakes, J. (1982). Classroom social relationships: Exploring the bowles and gintis hypothesis.

Sociology of Education, 55, 197-212.
Oakes, J. (1985). Keeping Track: How Schools Structure Inequality. New Haven, CT: Yale University Press.
Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science.* Santa Monica, CA: RAND.
Ochs, E., Taylor, C., Rudolph, D., & Smith, R. (1992). Storytelling as a theory-building activity. *Discourse Processes, 15*, 32-72.
Opdenakker, M.-C. & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematic achievement. *British Educational Research Journal, 27,* 407-432.
Organization for Economic Cooperation and Development [OECD]. (2000). *Education at a glance: OECD indicators*. Paris: OECD.
Ostrove, J. M., Feldman, P., & Adler, N. E. (1999). Relations among Socioeconomic Indicators and Health for African-Americans and Whites. *Journal of Health Psychology, 4*, 451-463.
Page, R. (1987). Lower-track classes at a college-preparatory school: A caricature of educational encounters. In G. Spindler & L. Spindler (Eds.), *Interpretive enthnography of eduation: At home and abroad (pp. 447-472)*. Hillsdale, NJ: Erlbaum.
Pallas, A., Entwisle, D. R., Alexander, K. M. & Stulka, P. M. (1994). Ability-group effects: instructional, social, or institutional. *Sociology of Education, 67,* 27-46.
Rasbash, J., & Woodhouse, G. (1995). *MLn command reference*. London: Multilevel Models Project, Institute of Education.
Reuman, D. A. (1989). How social comparison mediates the relation between ability-grouping practices and students' achievement expectancies in mathematics. *Journal of Educational Psychology, 81,* 178-189.
Rosenbaum, J. (1976). *Making inequality: The hidden curriculum of high school tracking*. New York: Wiley.
Rumberger, R. W., & Wilms, J. D. (1992). The impact of racial and ethnic segregation on the achievement gap in California high schools. *Educational Evaluation and Policy Analysis, 14*, 377-396.
Slavin, R. (1987a). Grouping for instruction: Equality and effectiveness. *Equality & Excellence, 23,* 31-36.
Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analysis and traditional reviews. *Educational Researcher, 75,* 5-11.
Slavin, R. E. (1987b). Ability grouping and student achievement in the elementary schools: A best-evidence synthesis. *Review of Educational Research, 57,* 293-336.
Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: *a* best-evidence synthesis. *Review of Educational Research, 60,* 471-499.
Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological Methodology* (Vol. 13, pp.290-312). San Francisco: Jossey-Bass.
Tieso, C. L. (2003). Ability grouping is not just tracking anymore. Roeper Review, 26(1), 29-36.
Van Houtte, M. (*2004*). Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education, 110,* 354-388.
Vanfossen, B. E., Jones, J. D. & Spade, J. Z. (1987). Curriculum tracking and status

maintenance. *Sociology of Education, 60,* 104-122. Webb & Palincsar, 1996

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal of Research in Mathematics Education, 22,* 366-389.

Willms, J. D. (1999). Quality and inequality in children's literacy: The effects of families, schools, and communities. In D. P. Keating & C. Hertzman (Eds.), *Developmental health and the wealth of nations* (pp. 72-93). New York: The Guilford Press.

World Bank. (2004). *The world development report 2003*. New York: Oxford University Press.