Within-School Tracking in South Korea

An Analysis Using Pisa 2003

Kevin Macdonald
Harry Anthony Patrinos

The World Bank
Human Development Network
Education Team
April 2010

Abstract

The 2003 PISA Korea sample is used to examine the association between within-school ability tracking and mathematics achievement. Estimates of a variety of econometric models reveal that tracking is positively associated with mathematics achievement among females and that this association declines for higher achieving

females. No evidence of an association between males and tracking is detected. While this association for females cannot be interpreted as a causal effect, the presence of a measurable association indicates the need for further research on tracking in Korea with a particular focus on gender differences.

This paper—a product of the Education Team, Human Development Network—is part of a larger effort in the department to analyze the determinants of learning. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at hpatrinos@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

WITHIN-SCHOOL TRACKING IN SOUTH KOREA AN ANALYSIS USING PISA 2003

Kevin Macdonald¹ and Harry Anthony Patrinos²

¹ Consultant, Human Development Network – Education Team, World Bank;
kadmacdonald1@worldbank.org

² Lead Education Economist, Human Development Network – Education Team, World Bank
(corresponding author) hpatrinos@worldbank.org

1. Introduction

Understanding the relationship between learning achievement and tracking is especially relevant to Korea due to its so-called equalization or leveling policy where, from 1974 onward, different areas began to assign students to upper secondary schools using a lottery system instead of by academic performance³ (Kim et al. 2006; Educators without Borders 2007). As a result, areas subject to equalization can be characterized as a mixed school system since students of different abilities attend the same schools; in the other areas, a tracked school system exists where students in each school are more homogenous in ability.

Some research has tried to measure the impact of this policy. For example, Kim et al. (2006) use a difference-in-differences method and find that the effect of the policy is negative. Kang et al. (2007) examine the impact of this policy on adult earnings and find positive effects for low ability students and smaller effects for high ability students.

Literature on tracking generally cites two paths through which tracking affects learning achievement (Kim et al. 2008; Duflo et al. 2008; Hanushek and Wößmann 2005). In the first, tracking benefits students because it creates a more ability-homogenous learning environment which allows teachers and schools to choose pedagogy better suited to a higher proportion of students. In the second, tracking inhibits low achieving students who would otherwise benefit from exposure to high achieving peers while their high achieving peers remain unaffected.

Consequently, recent empirical literature on tracking attempts to estimate the magnitude of peer effects. Some recent estimates, as presented in Ammermeuller and Pischke (2006), are small: an increase in peer composition of one standard deviation generally results in an increase in achievement between 0.05 to 0.11 standard deviations. The principle empirical challenge in this literature is distinguishing peer effects, the effect that a student's peers have on his or her achievement, from *correlated effects*, the effect that unobserved factors have on both the student and his or her peers. Correlated effects stem from a number of sources including an unobserved characteristic of the teacher or endogeneity in the selection of the peer group such as better students attending better schools. Several different strategies have been adopted to overcome this identification problem. For example, Ammermeuller and Pischke (2006) rely on random assignment of teachers and students to classes; Gibbons and Telhaj (2006) take advantage of random variation of primary students' assignments to secondary schools; Schneeweis and Winter-Ebmer (2005) use a fixed effects model to focus on peer effects within schools where class composition is generally random, and Fertig (2002) uses an instrumental variables approach.

Kang (2006) examines peer effects in the context of South Korea, also using random student assignment and instrumental variables. He finds higher ability peers have a causal and positive impact on the achievement of a student. Additionally, he finds that a low

³ Initially the policy's official name was the "Equalization of High School Policy" (Educators without Borders 2007).

achieving student tends to interact more closely with other low achieving students while a high achieving student tends to interact more closely with other high achieving students; this complicates the connection between peer effects and ability tracking: exposure to high achieving students may benefit a low achiever, but he or she would be hindered by exposure to his or her low achieving peers.

The proceeding analysis also provides a Korea-specific contribution by examining the 2003 Korea PISA sample. This paper examines the association between mathematics achievement and within-school tracking where schools assign students into groups depending on their ability. Particularly, this paper focuses on how this association differs between males and females.

In the following sections, estimates of linear regression models present the association between learning achievement and tracking and reveal a statistically significant association for females, no statistically significant difference for males, and a statistically significant difference in association between genders. Using PISA's categorical measure of achievement, its mathematics *proficiency levels*, multinomial logit estimates show, among females, a statistically significant decline in the association between tracking and the odds of being in one proficiency level relative to the previous for higher levels of proficiency. Finally, comparisons of the standard deviation of achievement between males and females in schools with tracking and not with tracking, as well as estimates of conditional quantile regressions, reveal no evidence that tracking associates with inequities in achievement.

However, none of the proceeding results should be interpreted as causal. The cross-sectional nature of the PISA dataset, in conjunction with unobserved factors, prevents establishing a counterfactual: we can not estimate the achievement of a student in a school with tracking had he or she been in a school without tracking. Furthermore, grade 10 students compose almost the entire sample, and the PISA examination occurred in June, 2003; since grade 10 is the first year of high school, students have only been in the observed school for three months; consequently, students have been exposed to tracking or non-tracking for a maximum of three months although likely less since midterm exams which are used to determine a student's ability are often held in late April. While we find statistically significant associations, they likely reflect other factors that are correlated with tracking but not observable. However, these associations are certainly interesting and expose a need for more in-depth, empirical research on tracking in Korea.

2. Data

The Programme for International Student Assessment (PISA) is the Organization for Economic Cooperation and Development's (OECD) international student assessment targeted to 15 year olds in grade 7 or higher, and it has been conducted every three years since 2000. In addition to testing students' ability in mathematics, science, and reading, PISA also collects background data about each student, his or her family, and his or her school. Schools are the primary sampling unit in PISA, and they are selected by one or more sampling frames or explicit strata. In Korea, there are several strata which divide the sample into general and vocational schools and by metropolitan, urban, and rural areas. For this paper, only observations from the general school strata and for grade 10 students were included.

Both the 2003 and 2006 PISA school questionnaires include questions about tracking within school; however, in 2003, the question specifies tracking among mathematics classes while in 2006, the questions specifies any subject. In order to clearly estimate the association between tracking and learning achievement, this paper uses the 2003 dataset and focuses on mathematics achievement. The school tracking question for 2003 is drawn from the principal's questionnaires:

Schools sometimes organize instruction differently for students with different abilities and interests in Mathematics. Which of the following options describe what your school does for 15-year-old students in Mathematics classes? (question 16)

The options included (a) Mathematics classes study similar content, but at different levels of difficulty; (b) Different classes study different content or sets of Mathematics topics that have different levels of difficulty; (c) Students are grouped by ability within their Mathematics classes; (d) In Mathematics classes, teachers use a pedagogy suitable for students with heterogeneous abilities (i.e. students are not grouped by ability). For each option, a principal could choose either "for all classes," "for some classes" or "not for any classes."

The three different types of tracking captured by parts (a) through (c) of question 16 all occur in Korea. The Ministry of Education encourages schools to divide and instruct students in three cohorts based on ability, although some schools divide students into two cohorts. Additionally, many schools will divide students by ability only for a portion of the weekly instruction and maintain mixed classes for the remainder in response to parents and students wanting homogeneous assessment of students. Table 1 outlines how school principals responded to the question:

Table 1: School Responses to Question 16

	Classes:	All	Some	None
a) Mathematics classes study similar content but at different levels of difficulty		0.1038 (0.0312)	0.5628 (0.0811)	0.3334 (0.0879)
b) Different classes study different content or sets of Mathematics topics that have different levels of difficulty		0.02 (0.01)	0.5206 (0.0796)	0.4595 (0.0809)
c) Students are grouped by ability within their Mathematics classes		0.0763 (0.0341)	0.66 (0.07)	0.2687 (0.0618)
d) In mathematics classes, teachers use a pedagogy suitable for <students with heterogeneous abilities>		0.1333 (0.0373)	0.7577 (0.052)	0.11 (0.03)

Source: 2003 PISA School Sample, General High School Strata

Note: Standard errors are reported in parentheses.

Given the varying types of tracking in Korea, part (c) seems to capture the notion of tracking which most closely corresponds to the current literature on tracking and peer effects. This question is the only one to emphasize ability as the method to divide students, but its interpretation could potentially be problematic: for example, one could reasonably interpret this as students within *each* mathematics class are grouped by ability or as students *among* mathematics class are grouped by ability. The variable *some tracking* will be a binary variable equal to one if a school's response to part (c) is "for some classes" or "for all classes" and zero if the response is "not for any classes". Note that this is a school level variable, and it does not tell us how or whether a particular student in the PISA sample may have been tracked.

Like other major student assessments such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), PISA treats a student's achievement as an unobservable random variable with a distribution conditioned on his or her performance on a standardized test as well as information about his or her background (OECD 2003). Consequently, PISA does not provide a single estimate of achievement, but rather, for each subject, five random draws from the latent variable's conditional distribution called plausible values. These random draws are then incorporated into the estimation of statistics which are functions of achievement such as means and regression coefficients.

Additionally, PISA categorizes ranges of achievement according to the typical proficiencies displayed by students in those ranges; these are called proficiency levels (OECD 2007). Mathematics and science achievement each are categorized into six proficiency levels while reading achievement is categorized into five. Since proficiency levels stem from achievement, for each subject, each student has five plausible proficiency levels.

3. Stochastic Model

We assume that the observations of the PISA's school, student and family background variables do not exactly determine the corresponding levels of student achievement, but rather that they give us more information about what these achievement levels could be. In this sense, the vector of achievements for the sampled group of students, y , is a random variable conditionally distributed by the set of background variable observations, matrix x , as expressed by the following stochastic model:

$$(1) \quad y | x \sim f(y, x)$$

Furthermore, we assume that the conditional mean of the distribution function f has the following property:

$$(2) \quad E[y | x] = x \beta$$

where β is a column vector of regression coefficients comprised of one parameter for each background variable in x . Given the setup of (1) and (2), each regression coefficient can be interpreted as the change in a student's expected achievement *associated* with a marginal change in the value of the coefficient's corresponding variable; since (2) is an assumption about the conditional distribution function, f , and not derived from a cognitive production function or any framework implying causation, the regression coefficients are not interpreted as marginal effects.

For analyzing how background variables associate to the dispersion of the conditional distribution, f , we assume that the 20th, 40th, 60th and 80th conditional quantiles of $y | x$ are linear functions of the background variables, x .

$$(3) \quad \text{prob}\{y_i \leq x_i \alpha_q\} = q \quad \forall q \in \{0.2, 0.4, \dots, 0.8\}, \forall i \in I$$

where I is the set of sampled students, y_i and x_i are student i 's achievement and background variable observations, respectively, and α_q is a column vector of q^{th} quantile regression coefficients with one parameter for each background variable. Analogous to the linear regression coefficients, β , each q^{th} quantile regression coefficient of α_q can be interpreted as the change in the q^{th} quantile of a student's distribution of achievement associated with a marginal change in the corresponding background variable. In other words, while the regular regression coefficients measure how variables associate with the mean of a student's distribution of possible achievements given his or her background variables, the q^{th} quantile regression coefficients measure how variables associate with the q^{th} quantile of a student's distribution of possible achievements given his or her background variables. Examining how these coefficients change for different quantiles describes how variables associate with the dispersion of the conditional distribution; a variable that is, for example, positively associated with the 80th quantile and negatively with the 20th quantile would be associated with an increase in this dispersion.

In order to use PISA's categorical measure of achievement, its proficiency levels, we assume that the probability of a particular student's proficiency level, p_i , is conditionally distributed on x_i by a multinomial logistic function:

$$(4) \quad \text{prob}\{p_i = j \mid x_i\} = \frac{\exp\{\varphi_j x_i\}}{1 + \sum_{k=0}^K \exp\{\varphi_k x_i\}} \quad \forall j \in \{2, \dots, K\}$$

where K is the number of proficiency levels and φ_j is the column vector of multinomial logit coefficients with one parameter for each background variable. (4) implies that

$$(4') \quad \ln \left\{ \frac{\text{prob}\{p_i = j \mid x_i\}}{\text{prob}\{p_i = j-1 \mid x_i\}} \right\} = (\varphi_j - \varphi_{j-1})x_i \quad \forall j \in \{2, \dots, K\}$$

or, in other words, that the association between a marginal change in any particular background variable and the log change in the odds of being in proficiency level j relative to $j-1$ is the corresponding component in the vector equal to $\varphi_j - \varphi_{j-1}$.

Estimating (4') is conceptually similar to dividing the sample into subsamples comprised of consecutive pairs of proficiency levels and then estimating logit models for each subsample. If the association between the odds of being in the next proficiency level and variables such as *some tracking* are stronger for lower achieving students than higher achieving students, estimations of (4') will capture this. Additionally, statistical tests can reveal whether these association change for higher achieving students versus low achieving students.

4. Variables

In order to focus as much as possible on the association between learning achievement and *some tracking*, the variation in achievement attributable to other variables needs to be accounted for. Typical analyses of tracking and peer effects assume cognitive production functions and use similar variables. Table 2 lists variables used in some of these recent studies which can be generally classified into three levels: student level variables which includes student personal characteristics as well as those of their families, class level variables which includes characteristics of their teachers, classrooms, and peers, and school level variables which includes characteristics of the school.

Table 2: Background Variables Used in Previous Studies

Study	Topic	Background Variables
Kim et al. (2008)	Tracking	gender, mother's education, father's education, mother is a housewife, school's teacher's average age, average experience, school size, private school, girls school, boys school
Ammermeuller et al. (2006)	Peer Effects	age, gender, parent's origin, language spoken at home, number of books at home, number of persons living in household, class size, teacher's gender, teacher's education, teacher's experience, community size, daily instruction hours, shortage of staff, teaching materials and buildings
Gibbons et al. (2006)	Peer Effects	gender, ethnicity, free-school-meal entitlement, special educational needs, age, year
Schneeweiss et al. (2005)	Peer Effects	gender, grade, single parent family, number of siblings, ethnic Austrian, mother's education, socioeconomic status, cultural communication, books at home, educational resources, parents unemployed, school size, total hours of instruction, urban school
Zimmer (2003)	Peer Effects in Tracking	gender, father's occupation, mother's occupation, father's education, mother's education, teacher's gender, teacher's experience, teacher's age, teacher's education, class size
Fertig, M (2002)	Peer Effects	gender, whether parents work, intact family, native student, parent born abroad, second generation, other languages at home, parents education, homework feedback, poor school conditions, regular tests, teacher shortages, discipline problems, student teacher ratio
Betts and Shkolnik (1998)	Tracking	student teacher ratio, teacher education, teacher experience, race

Similar variables compiled from the 2003 PISA dataset will be used in the proceeding analyses and are presented below.

Since this paper is particularly interested in tracking, we want to distinguish the association with learning achievement of tracking from that of other school-level characteristics whether observed or unobserved. While not all school characteristics are observable in PISA, the observed characteristics might reflect the unobserved characteristics in some way. Consequently, this analysis estimates the correlations between several candidate school variables and *some tracking* in order to identify which school variables to include in the models. The extent to which these variables act as proxies for unobserved school characteristics correlated with tracking determines how isolated the resulting estimates of association between *some tracking* and achievement is. Of the several candidate school variables in the PISA sample tested for correlation with *some tracking*, Table 3 presents those that were statistically significant. The candidate school variables that were tested were public versus private, grade range, proportion of funding from different sources (5 variables), school autonomy measures (12 variables), frequency of assessments, use of assessment data (8 variables), school size, school community size stratum (3 variables), student teacher ratio, and the importance of a student's academic record for admission.

Table 3: School Variables Correlated with Tracking

Variable	Desc. of Association with Tracking	Proportion of sample true (%)
Public	positive	48
autonomy – hiring teachers	negative	33
autonomy – teacher salary increases	negative	90
autonomy – course offering	positive	96
student teacher ratio	positive	-
located in metropolitan stratum relative to rural	positive	-

Source: Korea PISA 2003

The variables measuring autonomy over hiring teachers and teacher salary increases are positively correlated with each other and negatively correlated with tracking. Since autonomy over hiring teachers displays more variation, it can serve as a proxy for autonomy over salary increases and any other unobserved characteristics that both might reflect. Autonomy over course offering has very little variation since 96 percent of schools reported this characteristic; this variable was tested for correlation with achievement, and the results were not statistically significant; for this reason, it will be excluded from the models.

Additionally, several variables are of interest in the Korean context. Whether a particular school is subject to the equalization policy or not can not be observed in the 2003 PISA sample, but the importance a school places on a student's academic record for his or her admission can be observed. Consequently, this variable is included. Also, some schools are co-educational and some are strictly male or female. Schools report the number of female students and the number of male students which can be used to construct a binary variable of whether the school is co-ed or not. Table 4 lists all the variables used in this analysis with some descriptive statistics.

Table 4: Summary Statistics of Achievement and its Covariates

	Mean	Std Dev.	25th Quantile	75th Quantile
Math Achievement Full Sample	542.23 (3.24)	92.39 (2.14)	479.26 (3.74)	606.14 (4.22)
Math Achievement, General Strata, Grade 10	570.02 (4.01)	83.11 (2.55)	514.75 (4.64)	626.27 (4.48)
Some Tracking for Math Classes	0.7 (0.05)			
Public School	0.37 (0.04)			
Hiring Teachers Autonomy	0.38 (0.05)			
Located in the Metropolitan Stratum	0.52 (0.01)			
Located in the Urban Stratum	0.38 (0.01)			
Located in the Rural Stratum	0.1 (0.01)			
Student Teacher Ratio	16.81 (0.15)	1.85 (0.27)	16.14 (0.2)	17.9 (0.15)
Co-ed School	0.43 (0.05)			
Student Record Important for School Admission	0.5 (0.05)			
Female	0.38 (0.04)			
Hours of Instruction in Math	4.3 (0.07)	1.6 (0.15)	3.33 (0)	5 (0.42)
Mother's Education: None	0.03 (0)			
Mother's Education: Primary	0.07 (0)			
Mother's Education: Lower Sec.	0.16 (0.01)			
Mother's Education: Upper Sec.	0.67 (0.01)			
Mother's Education: Post Sec.	0.07 (0)			
0 - 10 books at home	0.03 (0)			
10 - 100 books at home	0.39 (0.01)			
100+ books at home	0.58 (0.01)			

Source: PISA 2003 Korea, General High School Strata, Grade 10 Students (first row includes all observations)

Note: Standard errors reported in parentheses.

5. Mathematics Achievement and Tracking

Estimation of our models is complicated by the plausible value estimates of mathematics achievement as well as the presence of intra-cluster correlation and other issues related to the complex survey design. The estimation methodology used in this paper adheres strictly to the methodology recommended by the OECD (2005) which provides unbiased estimators and standard errors assuming the other standard, requisite assumptions for each model are met.

Table 5 presents three estimates of the linear regression model of (2). The first includes only the *some tracking* variable; the estimate of this coefficient is equivalent to the difference in the mean achievement of students in schools with tracking versus those in schools without tracking. As can be seen, this estimate is positive, implying students at tracking schools have a higher level of achievement, but it is not statistically significant implying the difference could simply stem from sampling variation.

The second estimation includes only the school variables, and the third includes all other variables. Only with the inclusion of student and family characteristics does the positive difference between tracking students and non-tracking students become statistically significant.

Being female is negatively associated with mathematics achievement which is a typical result for many countries. Additionally, being located in an urban or metropolitan stratum instead of a rural stratum is positively associated with achievement, even when differences in student and background characteristics are controlled for. Having a mother's education less than upper secondary is negatively associated with achievement. Having a mother with more than an upper secondary education is also negatively associated with achievement; however, Table 4 reveals that this represents a small proportion of the sample. Finally owning books positively associates with achievement as well.

Table 5: Achievement Linear Model Estimates

	(1)	(2)	(3)
Some Tracking for Math Classes	6.29 (12.08)	9.08 (10.8)	17.96** (8.89)
Public School		-14.49 (9.59)	-11.51 (9.91)
Hiring Teachers Autonomy		-5.95 (8.99)	-5.43 (7.99)
Located in the Metropolitan Stratum		47.88*** (17.53)	26.24 (16.88)
Located in the Urban Stratum		42.83** (16.34)	29.83** (14.22)
Student Teacher Ratio		4.51* (2.49)	3.96* (2.21)
Co-ed School			-5.65 (11.61)
Student Record Important for School Admission			10.02 (7.77)
Female			-15.18*** (4.72)
Hours of Instruction in Math			1.11 (1.38)
Mother's Education: None (relative to Upper Sec.)			-7.85 (9.17)
Mother's Education: Primary (relative to Upper Sec.)			-11.01 (7.31)
Mother's Education: Lower Sec. (relative to Upper Sec.)			-10.86** (4.28)
Mother's Education: Post Sec. (relative to Upper Sec.)			-12.68** (6.17)
0 - 10 books at home (relative to 10-100)			-43.55*** (8.39)
100+ books at home (relative to 10-100)			36.16*** (3.71)
Constant	566.1*** (10.35)	455.06*** (36.85)	452.57*** (34.51)
Number of Observations	3524	3484	2793
Average R-Square	0.0012	0.0695	0.162

Note: Standard errors are reported in parentheses and significance at the 1%, 5%, 10% levels is denoted ***, **, and *, respectively. Estimates were calculated using 5 plausible values for estimates of achievement and Balanced Repeated Replication with Fays(0.5) adjustment were used to estimate the standard errors.

Source data: PISA 2003 Korea, General School Strata, Grade 10 Students

6. Gender Differences and Tracking

In order to examine whether the association between learning achievement and *some tracking* differs for males and females, Table 6 presents re-estimates of the linear model which include an interaction term for gender and *some tracking*.

Table 6: Achievement Linear Model Estimates with Female and Tracking Interaction

	(4)	(5)	(6)
Female	-21.17*** (4.96)	-48.44*** (9.75)	-36.21*** (7.96)
Some Tracking for Math Classes		-7.3 (10.91)	7.59 (8.67)
Female x Some Tracking		38.85*** (10.84)	30.05*** (10.49)
Other Controls	No	No	Yes
Constant	578.13*** (4.22)	583.31*** (9.05)	464.94*** (31.65)
Number of Observations	3642	3524	2739
Average R-Square	0.0154	0.0272	0.168
Female + Some Tracking x Female		-9.59* (5.27)	-6.16 (6.09)
Some Tracking + Some Tracking x Female		31.55** (14.91)	37.64*** (11.97)

Note: Standard errors are reported in parentheses and significance at the 1%, 5%, 10% levels is denoted ***, **, and *, respectively. Estimates were calculated using 5 plausible values for estimates of achievement and Balanced Repeated Replication with Fays(0.5) adjustment were used to estimate the standard errors.

Source data: PISA 2003 Korea, General School Strata, Grade 10 Students

The estimated model (4) includes only a binary variable for being female. By definition, the estimate for this coefficient is equivalent to the difference in the mean achievement of males and females, and the constant is equivalent to the mean math achievement of males. Average achievement for females is 21 points below the average achievement of males.

In model (5), an interaction variable and *some tracking* are added to the model. By definition, the difference between average achievement of females at tracking schools versus those not at tracking schools is the sum of the coefficients for *female* and *female x some tracking*. As indicated, females in tracking schools achieved, on average, 32 points higher than those not at tracking schools.

Also, by design of model (5), the sum of the parameters for *female x some tracking* and *female* is equivalent to the difference in mean achievement between males at tracking schools and females at tracking schools; whereas, the coefficient on *female* alone is the difference between males and females not at tracking schools. Since *female x some tracking* is positive, the gender gap in achievement is lower at tracking schools.

Finally, model (5)'s coefficient for *some tracking* is the difference between males in tracking schools versus those not in tracking schools. As can be seen, there is no significant difference. In other words, we observe a measurable difference in achievement between tracking and non-tracking schools only for females. Model (6) differs from model (5) with the inclusion of controls for other school, student, and family background variables. The conclusions drawn from this model are the same as drawn from (5), except, now they have taken into account differences in the other background variables. The chief conclusion of these results is that *some tracking* is positively associated with mathematics achievement for females; no evidence of this exists for males. Additionally, the association is different for males than females.

7. Low Achievers and Tracking

Low achieving students may benefit from tracking due to a homogenous learning environment or may be inhibited by tracking due to a lack of exposure to higher achieving peers. In the preceding estimations, a significant difference in math achievement associated with tracking is found only among females, and this difference is positive. While this result tends to be consistent with the connection between learning and ability-homogeneous peers, the preceding estimations do not relate this difference to achievement level; it is possible, for example, that tracking associates negatively for low achievers while positively for middle and higher achievers.

Previous studies which sought to measure peer effects among low achieving students typically have baseline data on the achievement of students in their sample (Duflo et al. 2008). For the PISA 2003 dataset, this is not the case. In order to use the PISA dataset to examine differences in tracking's association across the distribution of achievement, the multinomial logit model for proficiency levels of (4) is estimated. As shown in (4'), this model allows us to estimate the odds ratio associated with *some tracking* for different consecutive pairs of proficiency levels. This reflects whether this association is stronger or weaker for lower achieving students versus higher achieving students. Table 7 presents estimates of the proportion of general high school, grade 10 students at the various levels of mathematics proficiency. Table 7 reveals that most students are concentrated in levels 3 through 5.

Table 7: Proportion of Students at Each Mathematics Proficiency Level

Level 6	0.11 (0.01)
Level 5	0.22 (0.01)
Level 4	0.3 (0.01)
Level 3	0.22 (0.01)
Level 2	0.11 (0.01)
Level 1	0.03 (0.01)
Level <1	0.01 (0)

Note: Standard errors reported in parentheses.

Source: PISA 2003 Korea, General School Strata, Grade 10 Students

The standard errors of multinomial logit models increase with the number of categories in the dependent variable; consequently, this analysis groups several proficiency levels together. The number of categories used in this analysis is four: more than this creates very large standard errors and fewer decreases the variation in achievement.

The proceeding table presents estimates from two multinomial logit models where the categories below 4 are grouped as one and each proficiency level greater than or equal to 4 is assigned its own category. The subsequent results are roughly similar to merging proficiency levels 5 and 6 and creating an additional category for level 3. Each cell of the table presents the association between the covariate and the log of the odds for each pair of proficiency levels as well as the odds ratio in square brackets⁴.

The first estimated multinomial logit model, model (7), includes only the gender variable, the *some tracking* variable and their interaction. Analogous to the interpretation of model (5), the odds of a female being in level 4 relative to being in a level less than 4 is less than half (0.43 times) of that for males; however, for students attending a tracking school, it is not evident that females are any more or less likely than males to be in level 4 than in a lower level as indicated by the statistical insignificance of the sum of *female* and *female x some tracking*. Furthermore, among females, the odds of being in level 4 relative to being in a lower level for those in tracking is more than twice the odds for those females not in tracking as indicated by coefficient on the sum of *some tracking* and *female x some tracking*. The results are different, however, for higher proficiency levels. For example, it

⁴ Recall that an odds ratio is the change in odds which is equivalent to the change in relative probabilities. To calculate how many times the odds of achieving one proficiency level to that of a preceding level from the coefficient, b , calculate e^b (which is presented in square brackets).

is not evident that females are any more or less likely than males to be in level 6 than in level 5; among females, it is not evident that those attending a tracking school are any more or less likely to be in level 6 than in level 5. The main result is that tracking seems more strongly associated for lower achieving females than for higher achieving females.

Table 8: Mathematics Proficiency Level Multinomial Logit Model Estimates

Log of	(7a)	(7b)	(7c)	(8a)	(8b)	(8c)
	P{Level 4} P{Level <4}	P{Level 5} P{Level 4}	P{Level 6} P{Level 5}	P{Level 4} P{Level <4}	P{Level 5} P{Level 4}	P{Level 6} P{Level 5}
Some Tracking for Math Classes	0.13 (0.19) [1.14]	-0.23 (0.16) [0.79]	-0.24 (0.22) [0.79]	0.27 (0.18) [1.31]	-0.18 (0.21) [0.84]	0.2 (0.26) [1.22]
Female	-0.85*** (0.19) [0.43]	-0.42** (0.2) [0.66]	-0.07 (0.31) [0.93]	-0.72** (0.23) [0.49]	-0.42 (0.31) [0.66]	-0.35 (0.36) [0.70]
Female x Some Tracking	0.73*** (0.19) [2.08]	0.27 (0.21) [1.31]	-0.08 (0.34) [0.92]	0.65** (0.25) [1.91]	0.33 (0.32) [1.39]	0.16 (0.37) [1.17]
Other Controls	No			Yes		
Female + Some Tracking x Female	-0.12 (0.11) [0.89]	-0.15 (0.12) [0.86]	-0.14 (0.12) [0.87]	-0.07 (0.15) [0.93]	-0.08 (0.17) [0.45]	-0.19 (0.16) [0.83]
Some Tracking + Some Tracking x Female	0.87*** (0.2) [2.39]	0.05 (0.19) [1.05]	-0.32 (0.38) [0.73]	0.92*** (0.21) [2.51]	0.16 (0.23) [1.17]	0.36 (0.36) [1.43]

Note: Cells indicate the association between the row's denoted variable and the change in the log of the column's denoted odds ratio; these are calculated by using a linear combination of the standard coefficients estimated in the multinomial logit models. Odds ratios are reported in square brackets below. Standard errors are reported in parentheses and significance at the 1%, 5%, 10% levels is denoted ***, **, and *, respectively. Estimates were calculated using 5 plausible values for estimates of achievement and Balanced Repeated Replication with Fays(0.5) adjustment were used to estimate the standard errors.

Source data: PISA 2003 Korea, General School Strata, Grade 10 Students

Model (8) presents similar results when the control variables are included; although, the standard errors increase due to the higher number of variables.

From Table 8 alone, however, we can not conclude that this association *declines* for higher pairs of proficiency levels since the lack of significance for higher pairs may be due only to higher sampling variation and not to a weaker association. To test whether this association declines, table 9 presents statistical tests of the differences in the reported coefficients across the pairs of consecutive proficiency levels.

Table 9: Difference in Estimates Across Initial Proficiency Level

	from	(7 a) to (7 b)	(7 a) to (7 c)	(8a) to (8b)	(8a) to (8c)
Female + Some Tracking x Female		-0.03 (0.19)	-0.02 (0.15)	-0.01 (0.28)	-0.11 (0.21)
Some Tracking + Some Tracking x Female		-0.82*** (0.29)	-1.18*** (0.44)	-0.76** (0.35)	-0.55 (0.42)
Some Tracking		-0.36 (0.29)	-0.37 (0.26)	-0.44 (0.33)	-0.07 (0.26)
Female x Some Tracking		-0.46 (0.31)	-0.81** (0.39)	-0.32 (0.48)	-0.49 (0.37)

Note: Cells indicate the change in the estimated coefficients presented in the previous table; these are calculated by a using a linear combination of the standard coefficients estimated in the previous multinomial logit models. Standard errors are reported in parentheses. Significance at the 1%, 5%, 10% levels is denoted ***, **, and *, respectively.
Source: PISA 2003 Korea, General School Strata, Grade 10 Students.

The difference between females in tracking schools and not in tracking schools, as measured by the sum of *some tracking* and *some tracking x female*, is statistically different for low consecutive proficiency levels versus high ones. In other words, there is evidence that the association between tracking and learning achievement declines for females as their achievement increases.

For males, Table 8 reveals no statistically significant association between *some tracking* and the odds of being in the next proficiency level for any proficiency level, and table 9 does not reveal any change in this association.

The sum of the parameters for *female* and *some tracking* measures the log difference in the odds of being in the next proficiency level between males and females at tracking schools. This difference, however, is not statistically significant for any pair of proficiency levels nor is the change across pairs significant. The parameters for *female*, however, measure the difference between males and females who are not at schools with tracking; this difference is statistically significant only for low proficiency level pairs.

The interaction term, as presented in Table 9, exhibits a statistically significant change for higher proficiency level pairs; in other words, the gender difference in association of tracking declines for higher pairs.

8. Dispersion and Tracking

Besides learning outcomes, a major issue surrounding tracking is the possibility that it may reduce equity (Kim et al. 2008). Using the PISA dataset, we can examine the association between *some tracking* and the dispersion of achievement. Table 10 presents the standard deviation of achievement between students in schools with tracking and without for males and females. It also presents F statistics for comparisons across rows and columns; while the standard deviation is lower for both males and females in schools

with some tracking, the F statistics are too low for any of these differences be statistically significant.

Table 10: Standard Deviation Comparisons

	Tracking	No Tracking	F(80,80)
Males	81.77 (3.6)	88.03 (4.08)	1.077
Females	77.07 (3.64)	87.2 (7.94)	1.131
F(80,80)	1.061	1.01	

Note: Standard errors reported in parentheses.

Source: PISA 2003 Korea, Grade 10 Males, General School Strata

An alternative measure of how tracking associates to dispersion is to look at the conditional distribution of achievements. As modeled in assumption (1), for any particular student observation, his or her achievement is a random variable conditionally distributed by his or her background variables. In addition to the background variables that associate with this distribution's mean as in (2), these background variables may also associate with the shape of the dispersion. For example, the conditional distribution of achievement might be wider for students at tracking schools compared to those not at tracking schools if tracking schools reduce inequity. To estimate whether *some tracking* associates with the shape of the distribution of possible achievements for a particular student, the following table presents results from the quantile regression model of (3).

In addition to the results for the 20th, 40th, 60th and 80th conditional quantiles, estimates of the differences in coefficients between the 80th and 20th quantiles are also presented (Table 11). A negative difference for a variable means that the variable is associated with making a conditional distribution more compact while a positive difference reflects the opposite.

Table 11: Quantile Regression Estimates

	Quantile	0.2	0.4	0.6	0.8	diff. in coeffs 0.8 - 0.2
Female		-32.19** (12.63)	-38.47*** (12.09)	-40.42*** (13.73)	-41.89*** (12.44)	-9.71 (15.6)
Some Tracking for Math Classes		14.3 (11.94)	6.6 (10.25)	3.65 (9.59)	3.57 (9.24)	-10.73 (9.52)
Female x Some Tracking		22.6 (14.62)	27.27** (13.65)	29.76* (14.86)	28.88** (12.57)	6.29 (15.65)
Other Controls		Yes	Yes	Yes	Yes	
Constant		389.62*** (41.24)	455.98*** (39.08)	493.77*** (40.24)	570.57*** (37.21)	180.95*** (38.67)
Number of Observations		2970	2970	2970	2970	
Average R-Square		n/a	n/a	n/a	n/a	
Female + Some Tracking x Female		-9.59 (8.18)	-11.2 (6.82)	-10.66 (6.64)	-13.01* (6.53)	-3.42 (7.86)
Some Tracking + Some Tracking x Female		36.89*** (13.52)	33.87** (13.18)	33.41** (14.13)	32.45** (13.19)	-4.44 (13.39)

Note: Standard errors are reported in parentheses and significance at the 1%, 5%, 10% levels is denoted ***, **, and *, respectively. Estimates were calculated using 5 plausible values for estimates of achievement and Balanced Repeated Replication with Fays(0.5) adjustment were used to estimate the standard errors.
Source data: PISA 2003 Korea, General School Strata, Grade 10 Students

The interpretation of the coefficients and their sums is analogous to the interpretation presented in the previous sections except, instead of associating with the conditional mean, they associate with the indicated conditional quantile. For example, being female and not in a tracking school, as measured by the coefficient on *female*, is negatively associated with all four presented conditional quantiles. For females, being in a tracking school is positively associated with all four conditional quantiles as measured by the estimates of the sums of *some tracking* and *some tracking x female*. However, none of the differences between the 20th and 80th conditional quantile coefficients were statistically significant meaning we can not conclude that *some tracking* has any association with the shape of each student's conditional distribution. This is consistent with tracking not affecting equity.

One possibility is that tracking is actually related to school selection: suppose low ability female students prefer schools with tracking. In regions not subject to equalization, among lower ability female students, those with relatively higher ability would be more likely to gain admission to the school they prefer (with tracking) than those with relatively lower ability; high ability female students would be indifferent between choosing a school with tracking versus one without. The resulting pattern in achievement data would be similar to our results. This hypothesis is furthered by Table 12 which

compares math achievement levels among females between schools with tracking and without tracking for the rural, urban, and metropolitan strata.

Table 12: Mean Achievement of Girls in Tracking versus Non-Tracking Schools

	Metropolitan	Urban	Rural
Tracking	568.67 (6.58)	580.37 (11.2)	521.04 (16.69)
No Tracking	616.74 (39.33)	525.31 (15.03)	504.68 (14.31)
Difference	-48.07 (39.42)	55.06*** (18.96)	16.37 (18.33)

Note: Standard errors reported in parentheses.

Source: 2003 PISA Korea Sample, Female Students, General Highschool Strata, Grade 10 Students

While the limited number of observations prevents us from replicating the preceding analysis in only the urban stratum, the comparisons of mean mathematics achievement in Table 12 reveals a significant and positive difference between tracking and non-tracking schools only for female students in the urban stratum. In the metropolitan and rural strata, students have little choice as to which school they attend, either schools are subject to equalization and students are therefore assigned by lottery as in the metropolitan stratum or schools are too far away as in the rural stratum. But in many regions in the urban stratum, students are in regions whose schools are not subject to the equalization policy and as a result they do have a choice (Kim et al. 2008). Consequently, we cannot rule out that the difference in achievement associated with tracking might reflect selection stemming from a difference in preferences.

9. Conclusions and Future Research

In the 2003 PISA Korea sample, being in a tracking school is positively associated with mathematics achievement for female students, and, as shown by estimates of multinomial logit models, this association declines for higher levels of achievement. For males, however, there is no conclusive association.

But how do we interpret this association? It should not be interpreted as the “effect” on achievement of being in a school with tracking for two reasons: first, we can not establish the counterfactual levels of achievement for students who are in tracking had they not been, and second, students have only been exposed to their new schools for only three months which does not seem long enough for tracking to have an impact. More likely, the detected association between tracking and achievement stems from other factors. However, since we control for any school level variable that is correlated with tracking as well as the standard school, family, and student background variables, a lot of possible factors that these variables represent and proxy for, both observed and unobserved, can be ruled out.

In the Korea PISA data, girls' schools were no more or less likely to be a school with tracking than boys' schools or co-ed schools. Also, since we control for co-ed schools, it is unlikely that the measured association between achievement and tracking, as well as the gender difference in this association, reflects being at a girls school. Furthermore, schools where a student's academic record is important for admission (our proxy for schools subject to equalization) were no more likely to be a school with tracking than a school where academic record was not important to admission. As a result, it is unlikely the association between achievement and tracking reflects equalization.

Alternatively, being in a tracking school might have an immediate effect on mathematics achievement through a student's attitude or motivation towards mathematics. It is plausible that this effect would differ across the genders, and a follow-up analysis using PISA 2003 could gather more evidence relating to this.

Consequently, while we can not conclude from our analysis that tracking has an effect on student achievement, we do find a very interesting pattern in the 2003 dataset, and our results are especially interesting due to the gender difference aspect. Given the ongoing debate around the Korea's equalization policy, these results expose the need for more research on tracking in Korea with an emphasis on gender differences.

References

- Ammermueller, A. and J. Pischke (2006), "Peer Effects in European Primary Schools: Evidence from PIRLS", National Bureau of Economics R Working Paper 12180, Cambridge, MA.
- Betts, J. and J. Shkolnik (2000), "The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools", *Economics of Education Review* 19 1 - 15
- Duflo, E., P. Dupas, and M. Kremer (2008), "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya"
- Educators without Borders (2007), "Are Korean Kids Smart or Working Hard? Some Reasons behind Top-level achievers of Korean Students", Occasional Paper No. 1, Educators without Borders, Seoul, Korea
- Fertig, M (2003), "Educational Production, Endogenous Peer Group Formation and Class Composition - Evidence from the PISA 2000 Study", IZA Discussion Papers, 714, Bonn
- Gibbons, S. and S. Telhaj (2005), "Peer effects and pupil attainment: Evidence from secondary school transition", London School of Economics, mimeo.

Hanushek, E. and L. Wößmann (2005), “Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries”, *NBER*, Working Paper 11124.

Kang, C. (2007), “Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea”, *Journal of Urban Economics*, Vol 61, No. 3, pp 458 - 495

Kang C, C. Park, and M. Lee (2007), “Effects of Ability Mixing in High School on Adult Earnings: Quasi-Experimental Evidence from South Korea”, *Journal of Population Economics*, Vol. 20, No. 2, pp 269-297

Kim, T., J. Lee, and Y. Lee (2008), “Mixing versus sorting in schooling: Evidence from the equalization policy in South Korea”, *Economics of Education Review*

OECD (2003), *PISA 2003 Technical Report*, Paris: OECD

OECD (2005), *PISA 2003 Data Analysis Manual*, Paris: OECD

OECD (2007), *PISA 2006 Science Competencies for Tomorrow's World Volume 1: Analysis*, Paris: OECD

Schneeweiss, N and R. Winter-Ebmer (2005), “Peer Effects in Austrian Schools”, Department of Economics, University of Linz, Working Paper No. 0502

Zimmer, R (2003), “A new twist in the educational tracking debate”, *Economics of Education Review*, 22 307 – 315