Academic Interactions among Classroom Peers: A Cross-Country Comparison Using TIMSS

Changhui Kang
Department of Economics
National University of Singapore

August 2006
Applied Economics, forthcoming

Abstract

Using an international data set from the Third International Mathematics and Science Study (TIMSS), we examine academic interactions among classroom peers for each country, and compare them across different countries. To minimize the bias that usually plagues peer effects studies, we take within-student differences between mathematics and science test scores. The results show a significantly positive association between peers' performance and own achievement for most of the TIMSS countries. Moreover, the degree of mutual peer interactions within classroom is found to be surprisingly close across different countries, even if there exists a wide range of institutional differences in middle-school education (e.g. degree of ability mixing).

JEL *Classification* : I20, C20

Keywords : Peer Interactions, Ability Mixing, TIMSS

* Address: Department of Economics, National University of Singapore, 1 Arts Link, Singapore 117570, Singapore; E-mail: ecskch@nus.edu.sg, Phone: +65-6516-6830, Fax: +65-6775-2646.

Academic Interactions among Classroom Peers: A Cross-Country Comparison Using TIMSS

Abstract

Using an international data set from the Third International Mathematics and Science Study (TIMSS), we examine academic interactions among classroom peers for each country, and compare them across different countries. To minimize the bias that usually plagues peer effects studies, we take within-student differences between mathematics and science test scores. The results show a significantly positive association between peers' performance and own achievement for most of the TIMSS countries. Moreover, the degree of mutual peer interactions within classroom is found to be surprisingly close across different countries, even if there exists a wide range of institutional differences in middle-school education (e.g. degree of ability mixing).

JEL *Classification* : I20, C20
Keywords : Peer Interactions, Ability Mixing, TIMSS

1 Introduction

Peer group effects and social interactions receive growing attention from various perspectives. Studies report the presence of peer interactions for teen pregnancy and high-school completion (Evans et al., 1992), teenagers' criminal behavior (Glaeser et al., 1996; Ludwig et al., 2001), children's academic achievement and educational outcome (Aaronson, 1998; Hoxby, 2000*b*; Vandenberghe, 2002; Hanushek et al., 2003; Angrist and Lang, 2004; Cardak and McDonald, 2004), college students' grades, choice of major (Sacerdote, 2001; Zimmerman, 2003) and occupational choice (Marmaros and Sacerdote, 2002), students' drug/alcohol/substance use (Gaviria and Raphael, 2001; Kremer and Levy, 2003) and individuals' body weight (Costa-font, 2004).

Among various dimensions of peer interactions, the effect of classroom/school peers on a student's own academic achievement is at the heart of the debate on educational reform. Claims concerning ability grouping (Argys et al., 1996; Betts and Shkolnik, 2000; Figlio and Page, 2002; Kang et al., 2006), school desegregation (Angrist and Lang, 2004; Guryan, 2004), school choice (Epple and Romano, 1998) and school competition (Hoxby, 2000*a*; Hoxby, 2003; Epple et al., 2002) are in one way or another based upon a certain structure of academic interaction among peers. Nonetheless, the existence and nature of students' academic interactions remain controversial. Mostly from the US data, Hanushek (1971), Angrist and Lang (2004) and Arcidiacono and Nicholson (2005) find no significant peer achievement effects, while Hoxby (2000*b*), Zimmer and Toma (2000), Sacerdote (2001), Hanushek et al. (2003), Zimmerman (2003) and Winston and Zimmerman (2003) report significantly positive effects of peers' achievement on own academic outcome.

In spite of the potential importance of peers' academic interactions in educational production, detailed evidence on them is quite rare outside the US and Canada.[1] Instead of academic interactions among peers, a few existing studies outside these countries examine the effects of school/classroom peers' exogenous and contextual factors (e.g., socioeconomic status) on a student's academic outcome [Willms and Echols (1992) and Sawkins (2002) for Scotland; Glewwe (1997) for Philippines; Vandenberghe (2002) for OECD countries; McEwan (2003) for Chile]. In this paper we attempt to fill the gap by investigating academic interactions of classroom peers for countries outside as well as inside North America, and by comparing them across different countries. For that purpose, this study employs an international database of the Third International Mathematics and Science Study (TIMSS).

TIMSS is well suited for our examination on the similarity and differences of peer interactions among different countries. It covers a wide range of 41 different education systems and, for each system, it administers tests on two distinct subjects (mathematics and science) to two adjacent grades (usually grades 7 and 8). The rich country-specific information on a student and her family, teachers and school allows us to carry out an examination within a country and an extensive comparison between different countries.

Empirical detection of true peer interactions is often difficult due to various challenges. First, the selection problem plagues the empirical analysis. Students with similar personal and family backgrounds form a peer group and self-select into such a group. Such a sorting do not only take place across schools, but within a school. For example, in many countries schools educate different-ability students by means of tracked classrooms and curricula. The more tracked or stratified a classroom is in terms of ability, the more likely the estimate for peer interactions is to be (upward) biased. Second, a co-movement in outcomes of a student and her classmates or schoolmates may occur because both are subject to a common institutional environment. In addition, in a learning context, a student's behavior affects the behavior of her classmates and is reciprocally affected by them. Thus a student's outcome is simultaneously (or endogenously) determined with the outcome of peers—the reflection problem à la Manski (1993). To handle these empirical challenges, this study exploits the availability of two test scores (mathematics and science) per individual student in TIMSS. Specifically, it employs the fixed effects method of estimation and differences out the factors that may hamper an estimation of peer effects.

From the analysis we find the presence of academic interactions among classroom peers for most of the TIMSS countries—a significantly positive association between peers' performance and own achievement. Moreover, by pairing the estimate of peer interactions with a degree of ability mixing in education for each country, we show that the magnitude of peer interactions within classroom is surprisingly close across different countries, even if there exists a wide range of institutional differences in middle-school education. An additional merit of our cross-country comparison is to examine a potential bias that may arise to previous studies that use non-experimental data from countries adopting a high degree of ability grouping. Below we suggest that the conventional level-based specification overstates the estimates of peer interactions especially for those countries adopting a high degree of ability grouping within school.

The rest of the paper is organized in five sections. We review peer effects literature in Section 2. Section 3 specifies the empirical model and Section 4 describes the data. Section 5 presents the main estimation results. We conclude the paper in Section 6.

2 Academic Interactions among Peers

Peer effects in education illustrate the presence of a social interaction and an externality in generating an outcome. When there is a social interaction, a positive (negative) externality that arises from it expands (shrinks) the overall social effect beyond (below) the sum of individual marginal effects. In particular, social interactions in education have a great policy implication for designing the education system of a nation and class organization within a school (e.g. classroom tracking/ability-grouping versus untracking/ability-mixing). Studies—mostly from the US and Canada, however, report mixed results about the presence and magnitude of the effects of peer group on the individual student.

While Hanushek (1971) shows little effects of peer group quality on a student's academic outcome, Summers and Wolfe (1977) and Henderson et al. (1978) report significant effects of average peer quality on a student's achievement. These early studies are usually vulnerable to the sources of bias previously mentioned (Moffitt, 2001). To overcome the various types of limitations concerning peer-group formation, recent studies exploit either randomized (or semi-randomized) assignment of peers to individual students, or extremely comprehensive data sets which contain rich information on one's education background (e.g., the characteristics of school, teacher, peer groups and family, a history of one's growth in test scores, etc.).

Using random assignment of college roommates to a student, Sacerdote (2001) and Zimmerman (2003) show a significantly positive association between roommates' academic performance and a student's own achievement. In contrast, Stinebrickner and Stinebrickner (2004) report only limited evidence of a positive association based on similar random assignment in a college.

Employing a uniquely rich data set on Texas elementary school students, Hanushek et al. (2003) find strong influence of peers' academic quality (measured by previous academic performance) on own outcome. Hoxby (2000*b*) also relies on exogenous variation of peer groups in Texas elementary schools and finds significant positive effects of peers' achievement. On the contrary, examining students attending 124 US medical schools, Arcidiacono and Nicholson (2005) find no significant association between own and peers' achievements, once the school fixed effects are controlled for in the estimation.

There are a handful of non-US-based studies along this line. Employing an international data base (Second International Mathematics Study) from the International Association for the Evaluation of Educational Achievement (IEA), Zimmer and Toma (2000) show significantly positive effects of peers' quality on a student's own math test score for Belgium, France, New Zealand and Canada as a whole. Robertson and Symons (2003) find a positive association between peers' and own test scores for the UK. And, exploiting random assignment of peers taking place in middle schools of South Korea, Kang (2005) also reports a significantly positive association between own and peers' math-test scores.

The scarcity of empirical evidence on academic interactions among peers for various school systems may lead educational policy makers and researchers who are bound in a context of individual countries astray. To the extent that educational institutions may make differences in student academic achievement (Wößmann, 2003), current state of knowledge on peer interactions, which is largely based on North American and some European countries, leave much to be desired in search of country-specific educational policies. There will be little doubt that the information of peer interactions should be accumulated for various countries in order for the research to have better international relevance for policy issues.

3 Empirical Framework

For our empirical analysis we examine a model of educational production function given by:

$$y_{ij}^f = \beta_0 + P_i^f \beta_1 + Z_i \beta_2^f + X_i \beta_3^f + \gamma m_i + \alpha_i + \tau_j + u_{ij}^f \tag{1}$$

where y_{ij}^f is the standardized test score of student i in school j for math ($f = m$) and science ($f = s$); P_i^f is i's peer-group variables, that is, the mean and standard deviation of classroom peers' test scores (excluding i) of math (m) and science (s); Z_i is i's exogenous peer variables other than P_i^f (i.e. in-class proportion of boys, that of classmates who own computers and more than 200 books in the home, and in-class means of mother's and father's education of the classmates); m_i is an indicator variable for the math subject; X_i is the vector of i's personal and family backgrounds; α_i is i's unobservable attributes such as ability, motivation and parental involvement for learning; τ_j is an observable as well as an unobservable effect of school j; and u_{ij}^f is the random error term. Note that in its most comprehensive form, this specification allows for differential effects of observable personal and exogenous peer characteristics on math and science scores (i.e., $\beta_k^m \neq \beta_k^s$ for $k = 2, 3$), while assuming the same impact of unobservable person and school factors between the two subjects. We also suppose the effects of P_i^f are identical between math and science subjects for terseness of analysis.

It is well-known in the peer effects literature that P_i^f is highly correlated with τ_j. Students with a similar family background attend the same school due to parents' residential sorting by family income or parental education. And classmates/schoolmates share a common institutional environment in school. In addition, it is likely that even when the presence of association between P_i^f and τ_j is accounted for, P_i^f may still be correlated with α_i. Parents may pressure the school to appoint their child to a class with favorable peers, and the school independently groups (or tracks) students by ability level. The more tracked a classroom is, the more likely the estimate for peer interactions is to be biased. As a consequence, applying OLS to (1) yields biased results for peer interactions unless the two correlations between P_i^f, on the one hand, and α_i and τ_j, on the other, are somehow jointly considered.

In order to control for these non-zero correlations, we employ a fixed effects method. Specifically, exploiting the fact that both math and science test scores are available for each student from TIMSS, we take a within-student difference between the two subject scores.[2] The baseline (restricted) fixed effects model is given by

$$y_{ij}^{m} - y_{ij}^{s} = \gamma + (P_i^m - P_i^s)\beta_1 + (u_{ij}^m - u_{ij}^s) \tag{2}$$

Here we specify the same impacts of observable characteristics on math and science scores, that is, $\beta_k^m = \beta_k^s = \beta_k$ for all $k = 1, 2, 3$. To the extent that α_i and τ_j are differenced out, we obtain the estimates for peer interactions that is net of the problem of correlated unobservables and endogenous sorting to schools, on the one hand, and ability grouping within school, on the other.

Teachers are a common factor that is shared by all students in a classroom, hence P_i^f contains not only the interaction of classroom peers, but the influence of the teacher. To the extent that a good teacher raises student achievement, the addition of teacher variables to equation (1) and (2) may weaken the estimated effect of peer interactions. To address the teacher effects, we consider three types of observable teacher variables in the estimation, following studies that find the importance of teachers.[3] The first is a subject teacher's personal characteristics such as gender, academic qualification proxied by a master's degree and teaching experience (Rivkin et al., 2005). The second is a teacher's behavioral variables, which include the total number of weekly hours devoted to teaching-related activities (e.g., preparing or grading exams and other works, planning lessons, meeting with parents, and professional development), and the frequency of meeting with other teachers to discuss curriculum or teaching issues (Murnane and Phillips, 1981; Goldhaber and Brewer, 1997). Finally, studies show that matching the teacher's gender, race and/or ethnicity with students' matters in student achievement (Dee, 2004; Dee, 2005; Hanushek et al., 2005). The third variable is a dummy variable that indicates whether a subject teacher's gender matches with the student's.[4] When we control for the teacher variables, the sample size declines on average by 20 percent. Thus we report the estimate $\widehat{\beta}_1$ both with and without controls for the teacher variables. When the teacher variables are considered, we exclude the estimate of a country if its sample size diminishes by more than 30 percent.

Even if appropriate variables were controlled for in our estimation, the estimate $\widehat{\beta}_1$ would not reveal a causal relationship between peer quality and own achievement because of the 'reflection problem' (Manski, 1993). If there exist interactions among peers, a student's own outcome is affected by the performance of peers, which is reciprocally affected by her own behavior. This creates a simultaneous determination of own outcome and the outcome of peers, which implies that $E(P_i^f u_{ij}^f) \neq 0$ $(f = m, s)$. Therefore instead of showing the causal relationship, we interpret the estimate $\widehat{\beta}_1$ reveals the degree of mutual influences among students within classroom. When there are truly no interactions among classmates, we would not find $\widehat{\beta}_1$ to be significantly different from zero. $\widehat{\beta}_1$ being strictly non-zero is evidence suggesting the existence of peers' influences (Moffitt, 2001; Sacerdote, 2001).

As an extension of the baseline (restricted) model, we allow observable personal and exogenous peer characteristics of a student to have differential effects on math and science scores. Considering these differential impacts, the unrestricted model is expressed by

$$y_{ij}^m - y_{ij}^s = \gamma + (P_i^m - P_i^s)\beta_1 + Z_i\beta_2^m + X_i\beta_3^m + (u_{ij}^m - u_{ij}^s) \tag{3}$$

where it is assumed that $\beta_k^s = 0$ $(k = 2, 3)$ for identification.

For comparisons with the results of the difference-based models, we experiment with two level-based models of estimation. These two level-based models use a pooled sample of math and science scores, assuming that $\beta_k^m = \beta_k^s = \beta_k$ for all $k = 1, 2, 3$. First, we run OLS to (1), including only P_i^f, Z_i and X_i as explanatory variables and not explicitly controlling for the fixed effects of α_i and τ_j. The estimate for β_1 is known to be (probably upward) biased due to the presence of endogenous sorting into schools and ability grouping within school. Second, leaving α_i uncontrolled, we control for τ_j by school fixed effects. That is, we assign dummy variables for each school j to account for overall performance differences across schools that will mainly arise from across-school differences in available resources, quality of student bodies, neighborhood quality, etc. A usefulness of this specification is that after controlling for between-school differences in performance, we are able to examine the degree of bias in the estimate of peer interactions that can be attributed to ability grouping within school.

Different countries execute different methods of grouping (or mixing) students into schools and classrooms. For example, students and parents in the US have a fair amount of autonomy in choosing public schools—let alone private schools—within a school district, and schools often educate different-ability students through tracked classrooms and curricula (Argys et al., 1996; Betts and Shkolnik, 2000; Figlio and Page, 2002). In stark contrast, students and parents in South Korea enjoy little autonomy in choosing middle schools within a school district, and schools scarcely execute ability grouping: they assign different-ability students to a classroom nearly at random (Kang, 2005). The more mixed a classroom is in terms of ability, the more likely it is that a student's peers will be determined exogenously within school. Thus under the second level-based specification, we expect that countries adopting a high degree of ability grouping within school show larger (upward) bias in the estimate $\widehat{\beta}_1$ than those adopting a low degree of within-school ability grouping. Given that this second level-based specification is often used in studies of peer interactions (Hanushek et al., 2003; Betts and Zau, 2004; Lefgren, 2004; Vigdor and Nechyba, 2004), the analysis sheds light on potential bias that may contaminate empirical results of previous studies.

4 Description of the Data

For our empirical analysis, we employ data from the Third International Mathematics and Science Study (TIMSS), the tests conducted internationally by the International Association for the Evaluation of Educational Achievement (IEA) for 41 countries in 1994 and 1995.[5] For the countries participating in the study, mathematics and science tests were administered to three different populations toward the end of the school year : (1) Population 1: 9-year-old students, (2) Population 2: 13-year-old students, (3) Population 3: students in the final year of secondary education. In addition to the tests, TIMSS gathered detailed background information through separate student, teacher and school-principal questionnaires. This information is combined with an individual student's test outcomes. For this study, we focus on the mathematics and science test scores collected for the 13-year-old students who were attending middle schools (grades 7 and 8) at the time of the study.

For each TIMSS participating country, the rule was that at least 150 middle schools be selected (by a stratified sampling) to represent the national population of the relevant student group, and that two classes (one from grades 7 and 8 each) be sampled at random from each school. For most of the TIMSS countries, all students of each sampled classroom were tested.

However, there were a few exceptions —Colombia, Iran, Korea and Sweden. If not all students in a sampled classroom participated in the tests, tested students were randomly selected for testing and given appropriate weights to reflect the population of the sampled classroom.

There are three different weight variables in TIMSS to be assigned at each of the three levels of sampling: school, classroom and student.[6] As we are basically interested in what happens within a classroom, not in a school or a nation, we employ the unweighted raw class-level data for those countries in which all students of each sampled classroom were tested. For the countries in which a part of students were tested in the classroom, the raw class-level data are inflated to represent the classroom population by means of the integer part of [Student Weighting Factor(WGTFAC3)×Student Weighting Adjustment(WGTADJ3)+0.5] (see the footnote (6)). In addition, when it is necessary to obtain a national level reality (e.g., degree of ability mixing in a nation), we employ Total Student Weight (TOTWGT) constructed from school/class/student weighting factors. This weights an individual student with respect to a country's entire population.

As the two subject tests were administered to two different grades, we standardize the math and science scores using their national mean and standard deviation that are specific to the grade and subject. And, in order to maintain homogeneity of schools, we restrict our analysis to non-rural schools for each country.[7] The number of observations employed for the main analysis is given in the first column of Table 1.

INSERT TABLE 1 HERE.

5 Estimation Results

5.1 Estimate for Peer Interactions

The level-based OLS estimates and difference-based fixed effects estimates for β_1 are presented in Table 1.[8] Column (1) shows the level-based baseline estimate $\widehat{\beta}_1$ with no controls for α_i and τ_j, while column (2) shows the estimate with the control for τ_j alone, but not α_i. The difference-based restricted estimate $\widehat{\beta}_1$ under the restriction $\beta_k^m = \beta_k^s = \beta_k$ $(k = 1, 2, 3)$ is reported in column (4) and the unrestricted estimate in column (5). Columns (3) and (6) replicate the specification (2) and (5), respectively, with additional controls of the subject teacher's personal and behavioral variables and gender-matching.

From columns (1) and (2), we find that endogenous residential sorting associated with $E(P_i^f \tau_j) \neq 0$ yields inflated estimates of peer interactions. When both α_i and τ_j are not controlled for, the average $\widehat{\beta}_1$ across countries is 0.684 and its standard deviation is 0.147. When only τ_j is controlled for, $\widehat{\beta}_1$ falls for every country and the average becomes 0.336 (s.d. 0.256). Yet the estimate may be subject to bias for some countries due to the prevalence of ability grouping within school.

The difference-based estimates of β_1 also confirm the presence of bias in the estimates of columns (1) and (2). Compared with the level-based estimates in column (1), the size of difference-based estimates is small for all countries, whether we use restricted (in (4)) or unrestricted (in (5)) estimates for comparisons. The unrestricted difference-based estimates are 0.398 on average, their minimum is 0.072 from Kuwait, and the maximum is 0.663 from Korea. For most of the countries, the difference-based estimate $\widehat{\beta}_1$ is significantly different from zero at the 5 percent significance level. A similar comparison with column (2) reveals that the magnitude of difference-based estimates is usually smaller than the level-based estimate especially for countries adopting extensive ability grouping in education (e.g., Singapore, US, Hong Kong, Netherlands, etc.; see below).

It should be noted that the minimum and maximum of the unrestricted difference-based estimates of β_1 are within 2.03 standard-deviation range of the international average value (0.398). In other words, the estimates of β_1 are very close for most of TIMSS countries and indistinguishable from the international average value at 4.2 percent level of significance. Even if there exists a wide range of institutional and cultural differences in middle-school education, the degree of mutual peer interactions within the classroom appears surprisingly close across different countries. The average value implies that a one-standard-deviation increase in peers' mean test score is associated with a 0.4 standard-deviation higher own score. In terms of TIMSS raw test scores, a 100-point higher peers' mean score is associated with 40-point higher own score.[9]

5.2 Degree of Ability Mixing Across Countries

We indicate earlier that an estimate of peer interaction is subject to a bias when it is obtained by a level-based model with no controls for α_i and τ_j. It may still be subject to a bias when the level-based model considers τ_j alone, but not α_i. Under this second specification, we expect a smaller (larger) degree of bias when classrooms are mixed (grouped) in terms of students' ability in school. In order to examine this claim, we first need a measure for a country's degree of ability mixing in education.

Information, however, is very rare about the degree of within-school and across-school ability mixing across countries. Given this situation, we rely on two measures from TIMSS that indirectly reflects a country's degree of ability mixing in middle school. The first measure is the proportion of overall variance of standardized test scores that is attributed to within-class (or within-school) variance as opposed to between-class (or between-school) variance.[10] When ability mixing is widespread in a country, we expect a high fraction of within-class (or school) variance in total national variance of test scores. In contrast, when ability grouping is widely applied, the proportion of within-class (or school) variance will be relatively low, since similar-ability students will be gathered within a classroom (or a school). We calculate the proportions of such a within-variance separately for math and science. A country's index of the degree of ability mixing is the simple average of the two proportions. Table 1 shows two versions of this first index: one uses a classroom as a basic unit of measurement and the other a school. Both units produce similar figures. Correlation between the class-unit and school-unit measures of the degree of ability mixing is 0.954. We discuss the results on the basis of the class-unit measure of the degree of ability mixing below.

Our second measure is based on the school questionnaire that was administered to school principals to ask whether students followed the same course of study in mathematics. Different courses of study imply an ability grouping. We calculate the weighted (by Total Student Weight (TOTWGT)) proportion of students who were educated under the same course of mathematics as another index of ability mixing for each country. The correlation of the second index with the proportion of within-classroom variance is 0.64.

According to the first index, South Korea (0.963), Slovenia (0.927), Cyprus (0.925), Iceland (0.915), Kuwait (0.914), Norway (0.909), Iran (0.909) and Denmark (0.907) adopt high degrees of ability mixing in middle school. In contrast, Singapore (0.555), South Africa (0.565), Germany (0.586) and US (0.599) utilize high degrees of ability grouping. Other countries show an intermediate level of ability mixing in middle school. Cross-country trends of ability mixing are similar with respect to our second measure based on the same course of math study.

5.3 Degree of Ability Mixing and Bias in Estimates

Figure 1 plots the degree of class-unit ability mixing and the level-based estimates $\widehat{\beta}_1$ for each country whose values are presented in Table 1.[11] Table 2 shows the results of regressions whereby $\widehat{\beta}_1$ of each country is regressed against its degrees of ability mixing based on the class composition (in top panel), the school composition (in middle panel) and the same course of math study (in bottom panel). The column numbers in Table 2 correspond to those in Table 1 in terms of the specification that yields the estimate $\widehat{\beta}_1$.

INSERT TABLE 2 AND FIGURE 1 HERE.

Figure 1, and columns (1) and (2) of Table 2 reveal a significantly negative association between the degree of ability mixing and the level-based estimate $\widehat{\beta}_1$, whether the former is measured on the basis of the class composition, the school composition or the same course of math. This suggests that a country's degree of ability mixing in education is related with its size of the OLS estimate for peer interaction within classroom. To the extent that bias may arise from two major sources—correlations between P_i^f and τ_j, and between P_i^f and α_i, controlling for only τ_j in the level specification is informative of potential bias that can be attributed to the failure to consider the correlation between P_i^f and α_i. According to the right-hand plot of Figure 1, a bias still seems to exist for the countries adopting a relatively high degree of ability grouping in school, when the correlation between P_i^f and τ_j (but not α_i) is controlled for. A significantly negative association is shown between the degree of ability mixing and $\widehat{\beta}_1$.[12] A further comparison between columns (2) and (4) (or (5)) of Table 1 reveals that the level-based estimate are larger than corresponding difference-based estimates for countries adopting extensive ability grouping in education (e.g., Singapore, US, Hong Kong, Netherlands, etc.). This suggests that even if between-school differences are accounted for, a level-based specification may overstate peer interactions when within-school ability grouping is not properly considered. Data from countries adopting high degree of ability grouping would be more vulnerable to such a bias.

Figure 2 plots the degree of class-unit ability mixing and the difference-based estimate $\widehat{\beta}_1$ for each country. From columns (4) and (5) of Table 2 we find no systematic relationship between the degree of ability mixing and the estimate for peer interactions. Moreover, the unrestricted estimates are within the two standard-deviation range of the international average association.

That is, the degree of mutual peer interactions within classroom appears very similar across countries with different institutional backgrounds and degree of ability mixing. This also lends further support for the merits of the difference-based methods in equations (2) and (3).

INSERT FIGURE 2 HERE.

When teacher variables are added to the estimation models, few changes occur to $\widehat{\beta}_1$ for most of the countries. The average values of the level-based and difference-based estimates fall slightly. They become 0.293 and 0.369, respectively. According to Figure 3, a bias seems to exist in the level-based (school fixed effects) estimates of β_1 for some countries due to the prevalence of ability grouping within school. The difference-based (unrestricted) estimates imply that the degree of classroom peer interactions is still very close across countries and indistinguishable from the global average value.

INSERT FIGURE 3 HERE.

6 Concluding Remarks

In this paper we examine academic interactions among classroom peers for each TIMSS-participating country and compare them across different countries. We show that when the school fixed effects and a person's unobservable traits are not accounted for, the estimates for peer interactions are subject to an upward bias. Even if the school fixed effects are controlled for, the prevalence of ability grouping in education is shown to inflate the level-based estimates of peer interactions. Using two difference-based methods, we find a significantly positive association between peers' performance and own achievement for most of the TIMSS countries. Moreover, the degree of mutual peer interactions within classroom is shown to be surprisingly close across different countries, although there exist substantial differences in educational institutions, degree of ability mixing and cultural factors.

The presence of positive peer interactions illustrates the existence of externality in education. An externality suggests that the way in which students are grouped in a classroom or school can make a difference in the overall outcome of students. As Benabou (1996) presents, however, the efficient and outcome-maximizing method of class formation depends upon the degree of complementarity of a student's own ability and that of her peers. In order to assess the merit and weaknesses of ability mixing and grouping in education, we require a more detailed structure of peer interactions. That is, we need to know better whether students are more affected by weak peers or strong peers, and how the influences of weak and strong peers differ according to a student's ability levels. In this study we focus on potential effects of average peer quality on an *average* student's performance. Further exploration of peer interactions among different-ability students will shed light on the debates on ability grouping and mixing in education within and across countries.

Table 1: Estimates for Peer Interactions

		Degree of Mixing			OLS			Differences: Fixed Effects		
Country	N	Class	School	Same Math	(1)	(2)	(3)	(4)	(5)	(6)
Australia (AUS)	9,267	0.64	0.76	0.62	0.848	0.712	0.699	0.489	0.411	0.392
					(0.011)	(0.021)	(0.023)	(0.039)	(0.048)	(0.063)
Austria (AUT)[1)]	2,907	0.67	0.68	0.51	0.764	0.121	-0.007	0.444	0.333	0.342
					(0.026)	(0.105)	(0.105)	(0.071)	(0.079)	(0.091)
Belgium (BFL)[1)] (Flemish)	3,353	0.68	0.75	0.50	0.767	0.514	0.479	0.542	0.415	0.399
					(0.024)	(0.049)	(0.054)	(0.057)	(0.074)	(0.077)
Belgium (BFR)[1)] (French)	2,938	0.73	0.79	0.61	0.673	0.389	0.045	0.355	0.231	0.272
					(0.037)	(0.079)	(0.102)	(0.074)	(0.094)	(0.101)
Canada (CAN)	10,594	0.82	0.86	0.84	0.759	0.439	0.423	0.702	0.645	0.650
					(0.019)	(0.035)	(0.037)	(0.022)	(0.025)	(0.027)
Colombia (COL)	7,450	0.78	0.79	0.96	0.776	0.422	0.380	0.642	0.623	0.614
					(0.026)	(0.044)	(0.054)	(0.042)	(0.042)	(0.045)
Cyprus (CYP)	2,849	0.93	0.93	1.00	0.349	0.020	-0.049	0.230	0.231	0.071
					(0.079)	(0.124)	(0.120)	(0.101)	(0.102)	(0.165)
Czech Republic (CSK)	5,119	0.78	0.84	0.94	0.718	0.374	0.361	0.511	0.419	0.364
					(0.026)	(0.048)	(0.050)	(0.047)	(0.053)	(0.058)
Denmark (DNK)[1)]	1,238	0.91	0.92	0.97	0.467	-0.043	n.a.	0.465	0.421	n.a.
					(0.089)	(0.152)		(0.089)	(0.098)	
England (GBR)[2)]	2,666	0.79	0.79	0.37	0.568	n.a.	n.a.	0.109	0.096	n.a.
					(0.048)			(0.101)	(0.107)	
France (FRA)	3,043	0.81	0.84	0.72	0.701	0.259	0.232	0.444	0.367	0.375
					(0.030)	(0.074)	(0.078)	(0.062)	(0.067)	(0.079)
Germany (DEU)[1)]	2,698	0.59	0.60	0.79	0.769	0.358	n.a.	0.631	0.541	n.a.
					(0.023)	(0.06)		(0.038)	(0.064)	
Greece (GRC)	5,547	0.89	0.90	1.00	0.630	-0.092	-0.130	0.205	0.164	0.180
					(0.077)	(0.113)	(0.128)	(0.078)	(0.082)	(0.091)
Hong Kong (HKG)	5,562	0.63	0.67	1.00	0.904	0.631	0.623	0.593	0.490	0.465
					(0.009)	(0.046)	(0.048)	(0.059)	(0.076)	(0.087)
Hungary (HUN)	3,755	0.84	0.85	n.a.	0.496	0.048	0.053	0.286	0.225	0.209
					(0.047)	(0.088)	(0.091)	(0.066)	(0.072)	(0.081)
Iceland (ISL)	2,279	0.92	0.93	0.75	0.467	-0.211	-0.367	0.131	0.117	0.052
					(0.070)	(0.169)	(0.178)	(0.082)	(0.082)	(0.085)
Iran (IRN)[1)3)]	8,655	0.91	0.91	0.95	0.734	0.400	0.368	0.576	0.620	0.559
					(0.026)	(0.061)	(0.067)	(0.047)	(0.061)	(0.061)
Ireland (IRL)	4,386	0.66	0.77	0.21	0.778	0.603	0.600	0.482	0.425	0.370
					(0.021)	(0.043)	(0.046)	(0.056)	(0.063)	(0.079)
Israel (ISR)[4)6)]	1,256	0.82	0.83	0.57	0.620	n.a.	n.a.	0.523	0.503	n.a.
					(0.057)			(0.095)	(0.117)	
Japan (JPN)[5)]	8,216	0.90	0.90	1.00	n.a.	n.a.	n.a.	0.291	n.a.	n.a.
								(0.079)		
Korea (KOR)	13,533	0.96	0.96	1.00	0.707	0.494	0.467	0.727	0.663	0.637
					(0.022)	(0.039)	(0.038)	(0.027)	(0.035)	(0.040)
Kuwait (KWT)[4)6)]	1,576	0.91	0.91	n.a.	0.414	n.a.	n.a.	0.265	0.072	n.a.
					(0.085)			(0.137)	(0.160)	
Latvia (LVA)	2,462	0.90	0.93	1.00	0.590	0.237	0.231	0.488	0.487	0.506
					(0.038)	(0.084)	(0.081)	(0.064)	(0.069)	(0.069)
Lithuania (LTU)	3,308	0.75	0.81	0.83	0.761	0.518	0.521	0.442	0.392	0.405
					(0.024)	(0.057)	(0.058)	(0.070)	(0.081)	(0.092)
Netherlands (NLD)	2,185	0.63	0.67	0.39	0.768	0.418	0.398	0.406	0.280	0.217
					(0.021)	(0.078)	(0.083)	(0.082)	(0.095)	(0.111)
New Zealand (NZL)[6)]	5,334	0.69	0.70	0.65	0.758	n.a.	n.a.	0.468	0.432	0.360
					(0.023)			(0.056)	(0.059)	(0.067)

(Continued on the next page)

Table 1: Estimates for Peer Interactions (Continued)

		Degree of Mixing			OLS			Differences: Fixed Effects		
Country	N	Class	School	Same Math	(1)	(2)	(3)	(4)	(5)	(6)
							(Continued)			
Norway (NOR)[1)6)]	2,155	0.91	0.91	1.00	0.428	n.a.	n.a.	0.390	0.350	0.296
					(0.065)			(0.073)	(0.082)	(0.093)
Portugal (PRT)[1)]	5,579	0.89	0.91	0.98	0.534	0.107	0.062	0.280	0.186	0.170
					(0.039)	(0.068)	(0.074)	(0.061)	(0.076)	(0.081)
Romania (ROM)	5,217	0.71	0.76	0.98	0.859	0.624	0.614	0.656	0.621	0.556
					(0.017)	(0.040)	(0.041)	(0.05)	(0.051)	(0.047)
Russian (RUS)	5,928	0.73	0.81	0.90	0.853	0.625	n.a.	0.683	0.650	n.a.
Federation					(0.021)	(0.036)		(0.034)	(0.042)	
Scotland (SCO)	3,802	0.81	0.85	0.73	0.610	0.327	n.a.	0.395	0.393	n.a.
					(0.041)	(0.087)		(0.067)	(0.068)	
Singapore (SGP)	8,229	0.56	0.65	0.20	0.879	0.720	0.722	0.342	0.309	0.310
					(0.009)	(0.030)	(0.030)	(0.058)	(0.064)	(0.066)
Slovak Republic	4,912	0.82	0.87	0.79	0.704	0.358	0.334	0.450	0.404	0.393
(SLV)					(0.033)	(0.051)	(0.056)	(0.050)	(0.061)	(0.064)
Slovenia (SVN)[1)]	3,229	0.93	0.94	0.91	0.543	0.054	0.027	0.357	0.312	0.306
					(0.047)	(0.098)	(0.104)	(0.084)	(0.093)	(0.093)
South Africa	7,926	0.57	0.57	n.a.	0.882	n.a.	n.a.	0.386	0.364	n.a.
(ZAF)[4)6)]					(0.013)			(0.062)	(0.068)	
Spain (ESP)	3,807	0.87	0.88	1.00	0.567	0.017	-0.068	0.482	0.470	0.380
					(0.042)	(0.099)	(0.111)	(0.064)	(0.065)	(0.073)
Sweden (SWE)	4,988	0.84	0.90	0.42	0.579	0.137	0.112	0.312	0.276	0.284
					(0.031)	(0.088)	(0.089)	(0.058)	(0.062)	(0.068)
Switzerland	4,775	0.67	0.68	0.67	0.768	0.209	0.162	0.539	0.508	0.463
(CHE)					(0.018)	(0.071)	(0.080)	(0.050)	(0.051)	(0.060)
Thailand (THA)	3,452	0.74	0.77	0.77	0.878	0.601	0.552	0.723	0.608	0.397
					(0.016)	(0.051)	(0.062)	(0.043)	(0.070)	(0.120)
United States	7,420	0.60	0.73	0.14	0.809	0.683	0.666	0.546	0.486	0.439
(USA)					(0.014)	(0.028)	(0.032)	(0.043)	(0.043)	(0.059)
Mean:		0.78	0.81	0.75	0.684	0.336	0.293	0.450	0.398	0.369
S.D.:		0.12	0.10	0.26	0.147	0.256	0.285	0.157	0.161	0.150
Max:		0.96	0.96	1.00	0.904	0.720	0.722	0.727	0.663	0.650
Min:		0.56	0.57	0.14	0.349	-0.211	-0.367	0.109	0.072	0.052
Teacher Characteristics					No	No	Yes	No	No	Yes
School Fixed Effect (τ_j)					No	Yes	Yes	Yes	Yes	Yes
Student Fixed Effect (α_i)					No	No	No	Yes	Yes	Yes

Note: Robust standard errors are in parentheses.
1) Teacher's academic qualification (master's degree) is not controlled for due to low response.
2) Father's and Mother's education levels are missing and not controlled for in the regressions.
3) Age is not controlled for in the regressions due to the low response rate.
4) Rural schools are included due to the absence of the school location information.
5) Family-related information is missing.
6) Only one classroom is sampled per school.

Table 2: Regression of the Estimates for Peer Interactions on the Degree of Ability Mixing

Explanatory Variables	OLS (1)	OLS (2)	OLS (3)	Differences: Fixed Effects (4)	Differences: Fixed Effects (5)	Differences: Fixed Effects (6)
	A. Unit of Ability Mixing: Class					
Degree of Mixing	-0.989*	-1.491*	-1.584*	-0.388	-0.262	-0.187
	(0.137)	(0.334)	(0.399)	(0.213)	(0.219)	(0.261)
Intercept	1.451*	1.490*	1.523*	0.752*	0.602*	0.514*
	(0.099)	(0.254)	(0.303)	(0.160)	(0.161)	(0.188)
R-square	0.620	0.456	0.410	0.084	0.037	0.021
Number of Countries	39	33	29	40	39	31
	B. Unit of Ability Mixing: School					
Degree of Mixing	-1.045*	-1.439*	-1.612*	-0.401	-0.273	-0.239
	(0.160)	(0.452)	(0.567)	(0.239)	(0.236)	(0.312)
Intercept	1.532*	1.510*	1.613*	0.776*	0.620*	0.564*
	(0.126)	(0.374)	(0.469)	(0.190)	(0.184)	(0.242)
R-square	0.537	0.306	0.285	0.069	0.031	0.023
Number of Countries	39	33	29	40	39	31
	C. Unit of Ability Mixing: Same Course of Math Study					
Degree of Mixing	-0.163*	-0.358*	-0.345*	0.100	0.156	0.081
	(0.081)	(0.142)	(0.163)	(0.098)	(0.097)	(0.089)
Intercept	0.812*	0.614*	0.557*	0.386*	0.297*	0.314*
	(0.054)	(0.105)	(0.125)	(0.073)	(0.071)	(0.055)
R-square	0.091	0.133	0.107	0.027	0.068	0.021
Number of Countries	36	32	28	37	36	30

Note: Robust standard errors are in parentheses. * means the estimate is significant at the 5 percent level.

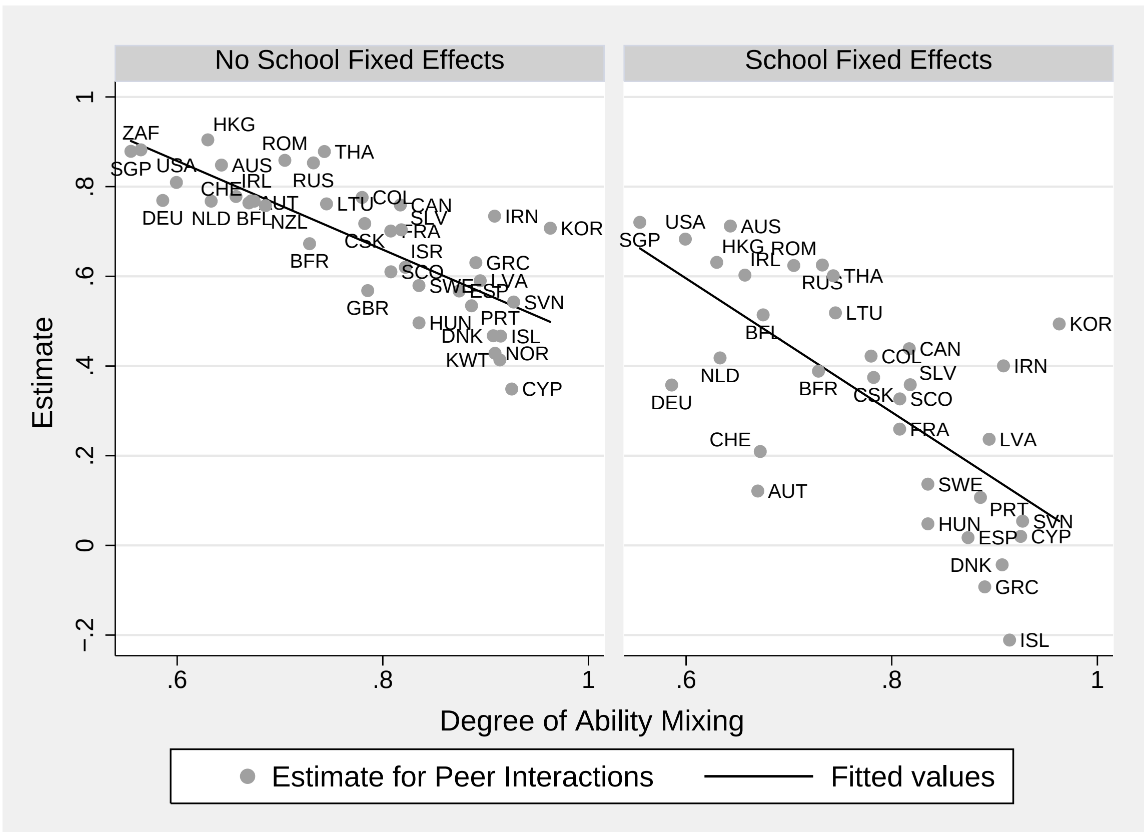

Figure 1: Level-based Estimates for Peer Interactions and Degree of Class-unit Ability Mixing

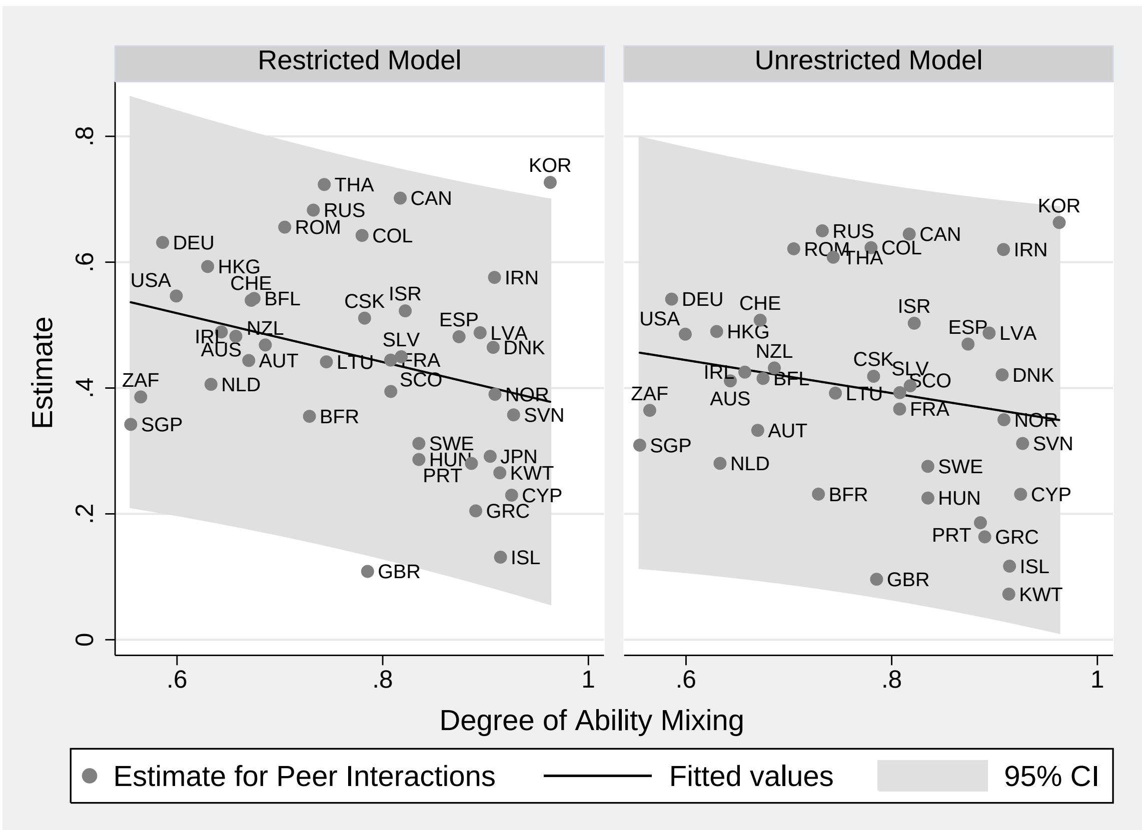

Figure 2: Difference-based Estimates for Peer Interactions and Degree of Class-unit Ability Mixing

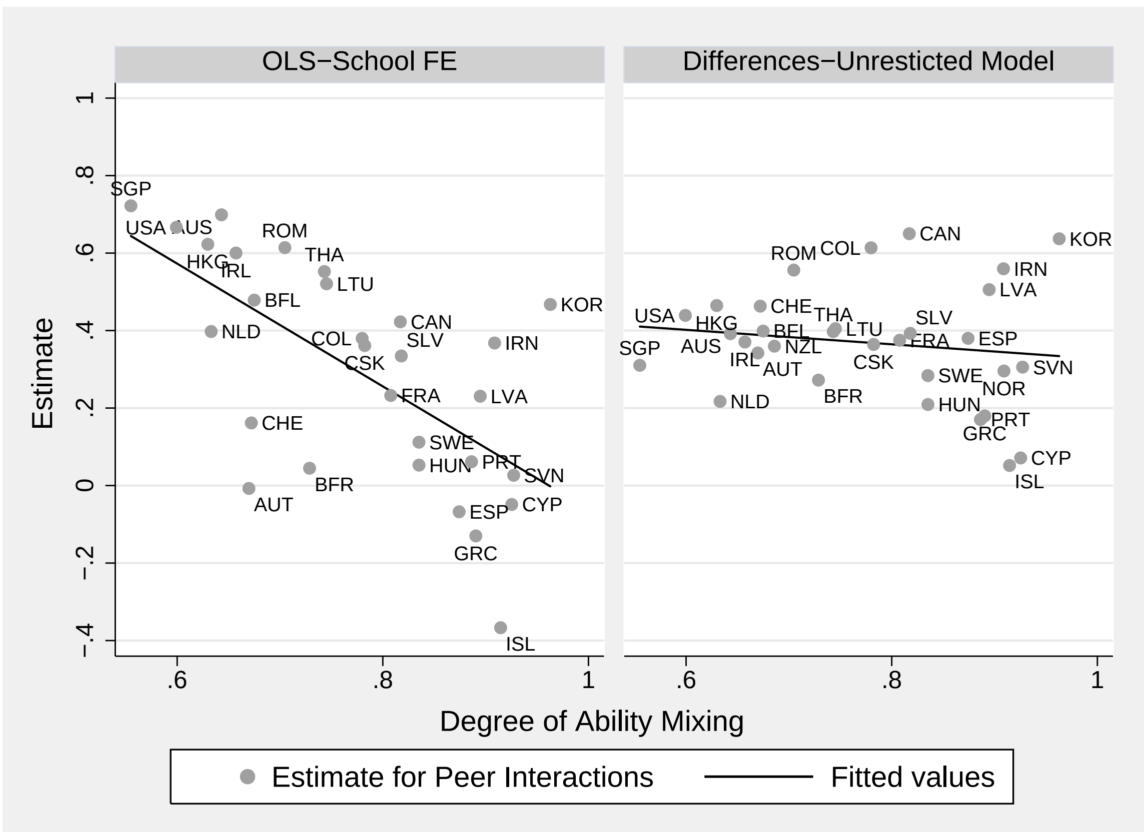

Figure 3: Estimates for Peer Interactions and Degree of Class-unit Ability Mixing: With Controls of Teacher Variables

Notes

[1]There are some exceptions. Academic interactions among classroom peers have been examined by Zimmer and Toma (2000) for Belgium, France, New Zealand and Canada as a whole, Robertson and Symons (2003) for the UK, and Kang (2005) for South Korea.

[2]In an estimation of the effect of neighborhood characteristics on children's educational outcomes, Aaronson (1998) employs a similar method of taking differences between siblings in order to control for the family's endogenous decision about a residential neighborhood.

[3]Controlling for only observable characteristics and behavior of teachers may not be sufficient to remove a teacher's influences. Recent studies document that unobservable teacher variables matter in student performance (Hanushek et al., 2003; Rivkin et al., 2005). An investigation of a teacher's effect in an unobservable dimension, however, usually requires longitudinal information for both student and teacher dimensions, which is not often available in data (Rockoff, 2004; Rivkin et al., 2005). Unfortunately, the limited information of the cross-sectional TIMSS data prevents us from examining the effect of unobservable teacher variables. Such an examination would be a topic of future research. We thank a referee for pointing out the need to control for teacher influences.

[4]Variables for race- and ethnicity-matching are not used in the absence of relevant information in the TIMSS data. In our analysis, if there are more than one teacher for a student, the variable for gender-matching represents the proportion of such matching among multiple teachers.

[5]See Martin and Kelly (1996) and Gonzalez and Smith (1997) for details of the database.

[6]The school-level weight variable is School Weighting Factor (WGTFAC1) and its adjustment (WGTADJ1). They are multiplied to produce the sampling weight for the school. The class-level weight variable is the Class Weighting Factor (WGTFAC2), which reflects the selection probability of the classroom within the school. The student-level weight variable is Student Weighting Factor (WGTFAC3) and its adjustment (WGTADJ3), whose product shows the selection probability of the individual student within a classroom. Obtained from these weight variables is Total Student Weight (TOTWGT), which shows the sampling weight of an individual student in a country's entire population.

[7]There are a few exceptions to this restriction. In the absence of school-related information, rural schools are included in the sample of Israel, Kuwait and South Africa. In addition to the current estimation excluding rural schools, we have estimated the same models, including rural schools for each country. When they are included in the analysis, the estimate for peer interactions for each country is not substantially affected, while some countries see it rising and others falling. Nonetheless, these changes do not affect the main results reported in section 5. The estimation results including rural schools are available upon request.

[8]Here we present only the estimates for mean peer scores. The estimates for the standard deviation of peer scores—a measure of classroom peer heterogeneity—are suppressed. They are available upon request.

[9]This amount of peer interaction is comparable to studies using US elementary schools. Hoxby (2000*b*) presents the 0.1 to 0.55-point increase in own score in association with a 1-point increase in peers' mean

score. Hanushek et al. (2003, Table I) show about 0.4-point increase in own math score in relation to a 1-point increase in peers' mean score. Vigdor and Nechyba (2004, Table 6) report a 1-point increase in peers' mean score is associated with 0.07-point increase in own math score.

[10]We decompose the total weighted (by Total Student Weight (TOTWGT)) variance of subject test scores into the within- and between-classroom (or school) variances for each country, as follows:

$$\sigma^2 = \sum_j F_j \sigma_j^2 + \sum_j F_j (m_j - \bar{m})^2$$

where σ^2 is the overall variance of math (or science) test scores, F_j is the fraction of students in classroom (or school) j, m_j and σ_j^2 are the mean and variance, respectively, of the test score within j, and $\bar{m}$ is the overall mean. The proportion of the within-variance is given by the ratio of $\sum_j F_j \sigma_j^2$ to σ^2. Here we employ the non-rural sample of each country, and the statistics are weighted by Total Student Weight(TOTWGT) in order to reflect the reality of a country.

[11]The pattern is similar when the degree of school-unit ability mixing or the proportion of students under the same course of math is employed. Such plots are available upon request.

[12]Boozer and Cacciola (2001) show a positive relationship between the class size and the OLS estimate of peers' mean outcome obtained from the canonical 'y on $\bar{y}$' specification. According to them, the smaller $\widehat{\beta}_1$ may mechanically result from the smaller class size. To address this possibility, we run the same regression as above, adding the average class size of each country as an explanatory variable. This does not alter the main results of the negative association between the degree of ability mixing and $\widehat{\beta}_1$, while the class-size coefficient is positive and significant. The estimate for mean peer scores under the new specification is -1.619 (s.e. 0.235) and that for the class size is 0.013 (s.e. 0.002).

References

Aaronson, D. (1998), 'Using sibling data to estimate the impact of neighborhoods on children's educational outcomes,' *Journal of Human Resources* **33**(4), 915–946.

Angrist, J. D. and K. Lang (2004), 'Does school integration generate peer effects? Evidence from Boston's Metco program,' *American Economic Review* **94**(5), 1613–1634.

Arcidiacono, P. and S. Nicholson (2005), 'Peer effects in medical school,' *Journal of Public Economics* **89**, 327–350.

Argys, L. M., D. I. Rees and D. J. Brewer (1996), 'Detracking America's schools: Equity at zero cost?,' *Journal of Policy Analysis and Management* **15**(4), 623–645.

Benabou, R. (1996), 'Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance,' *American Economic Review* **86**(3), 584–609.

Betts, J. R. and J. L. Shkolnik (2000), 'The effects of ability grouping on student achievement and resource allocation in secondary schools,' *Economics of Education Review* **19**(1), 1–15.

Betts, J. R. and A. Zau (2004), 'Peer groups and academic achievement: Panel evidence from administrative data,' *Unpublished Manuscript*.

Boozer, M. A. and S. Cacciola (2001), 'Inside the 'black box' of Project STAR: Estimation of peer effects using experimental data,' *Economic Growth Center Discussion Paper* **832**, Yale University.

Cardak, B. A. and J. T. McDonald (2004), 'Neighbourhood effects, preference heterogeneity and immigrant educational attainment,' *Applied Economics* **36**(6), 559 – 572.

Costa-font, Joan and Joan Gil (2004), 'Social interactions and the contemporaneous determinants of individuals' weight,' *Applied Economics* **36**, 2253–2263.

Dee, Thomas S. (2004), 'Teachers, race and student achievement in a randomized experiment,' *Review of Economics and Statistics* **86**(1), 195–210.

Dee, Thomas S. (2005), 'A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?' *American Economic Review* **95**(2), 158–165.

Epple, D., E. Newlon and R. Romano (2002), 'Ability tracking, school competition, and the distribution of educational benefits,' *Journal of Public Economics* **83**, 1–48.

Epple, D. and R. E. Romano (1998), 'Competition between private and public schools, vouchers, and peer-group effects,' *American Economic Review* **88**(1), 33–62.

Evans, W. N., W. E. Oates and R. M. Schwab (1992), 'Measuring peer group effects: A study of teenage behavior,' *Journal of Political Economy* **100**(5), 966–991.

Figlio, D. N. and M. E. Page (2002), 'School choice and the distributional effects of ability tracking: Does separation increase equality?,' *Journal of Urban Economics* **51**, 497–514.

Gaviria, A. and S. Raphael (2001), 'School-based peer effects and juvenile behavior,' *Review of Economics and Statistics* **83**(2), 257–268.

Glaeser, E. L., B. Sacerdote and J. A. Scheinkman (1996), 'Crime and social interactions,' *Quarterly Journal of Economics* **111** (2), 507–548.

Glewwe, P. W. (1997), 'Estimating the impact of peer group effects on socioeconomic outcomes: Does the distribution of peer group characteristics matter?,' *Economics of Education Review* **16**, 39–43.

Goldhaber, Dan D. and Dominic J. Brewer (1997), 'Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity,' *Journal of Human Resources* **32**(3), 505–523.

Gonzalez, E. J. and T. A. Smith (1997), *User Guide for the TIMSS International Database: Primary and Middle School Years*, TIMSS International Study Center, Boston College, MA.

Guryan, J. (2004), 'Desegregation and black dropout rates,' *American Economic Review* **94**(4), 919–943.

Hanushek, E. A. (1971), 'Teacher characteristics and gains in student achievement: Estimation using micro data,' *American Economic Review* **61** (2), 280–288.

Hanushek, E. A. (2003), 'The failure of input-based schooling policies,' *Economic Journal* **113**, F64–F98.

Hanushek, E. A., J. F. Kain, J. M. Markman and S. G. Rivkin (2003), 'Does peer ability affect student achievement?' *Journal of Applied Econometrics* **18**(5), 527–544.

Hanushek, E. A., J. F. Kain, D. M. O'Brien and S. G. Rivkin (2005), 'The market for teacher quality,' *NBER Working Paper* **11154**.

Henderson, V., P. Mieszkowski and Y. Sauvageau (1978), 'Peer group effects and educational production functions,' *Journal of Public Economics* **10**, 97–106.

Hoxby, C. M. (2000*a*), 'Does competition among public schools benefit students and taypayers?,' *American Economic Review* **90**, 1209–1238.

Hoxby, C. M. (2000*b*), 'Peer effects in the classroom: Learning from gender and race variation,' *NBER Working Paper* **7867**.

Hoxby, C. M. (2003), 'School choice and school competition: Evidence from the United States,' *Swedish Economic Policy Review* **10**(2), 11–67.

Kang, C. (2005), 'Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea,' *Working Paper.* Department of Economics, National University of Singapore.

Kang, C., C. Park and M.-J. Lee (2006), 'Effects of ability mixing in high school on adulthood earnings: Quasi-experimental evidence from South Korea,' *Journal of Population Economics* **forthcoming**.

Kremer, M. and D. M. Levy (2003), 'Peer effects and alcohol use among college students,' *NBER Working Paper* **9876**.

Lefgren, L. (2004), 'Educational peer effects and the Chicago public schools,' *Journal of Urban Economics* **56**(2), 169–191.

Ludwig, J., G. J. Duncan and P. Hirschfield (2001), 'Urban poverty and juvenile crime: Evidence from a randomized housing mobility experiment,' *Quarterly Journal of Economics* **116**(2), 655–679.

Manski, C. (1993), 'Identification of endogenous social effects: The reflection problem,' *Review of Economic Studies* **60**, 531–542.

Marmaros, D. and B. Sacerdote (2002), 'Peer and social networks in job search,' *European Economic Review* **46**, 870–879.

Martin, M. O. and D. L. Kelly (1996), TIMSS technical report Volume I: Design and development, Technical report, International Association for the Evaluation of Educational Achievement, Boston College, Chestnut Hill, MA.

McEwan, P. J. (2003), 'Peer effects on student achievement: Evidence from Chile,' *Economics of Education Review* **22**(2), 131–141.

Moffitt, R. A. (2001), Policy interventions, low-level equilibria, and social interactions, *in* S. N. Durlauf and H. P. Young, eds, 'Social Dynamics,' MIT Press: Cambridge, MA, pp. 45–82.

Murnane, Richard J. and Barbara Phillips (1981), 'What do effective teachers of inner-city children have in common?' *Social Science Research* **10**, 83–100.

Rivkin, Steven G., Eric A. Hanushek and John F. Kain (2005), 'Teachers, schools, and academic achievement,' *Econometrica* **73**(2), 417–458.

Robertson, D. and J. Symons (2003), 'Do peer groups matter? Peer group versus schooling effects on academic attainment,' *Economica* **70**, 31–53.

Rockoff, Jonah E. (2004), 'The impact of individual teachers on student achievement: Evidence from panel data' *American Economic Review* **94**(2), 247–252.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for Dartmouth roommates,' *Quarterly Journal of Economics* **116**(2), 681–704.

Sawkins, John W. (2002), 'Examination performance in Scottish secondary schools: An ordered logit approach,' *Applied Economics* **34**, 2031–2041.

Stinebrickner, T. R. and R. Stinebrickner (2004), 'What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds,' *Unpublished Manuscript.*

Summers, A. A. and B. L. Wolfe (1977), 'Do schools make a difference?,' *American Economic Review* **67**(4), 639–652.

Vandenberghe, V. (2002), 'Evaluating the magnitude and the stakes of peer effects analysing science and math achievement across OECD,' *Applied Economics* **34**(10), 1283–1290.

Vigdor, J. and T. Nechyba (2004), 'Peer effects in elementary school: Learning from "apparent" random assignment,' *Unpublished Manuscript.*

Willms, J. D. and F. Echols (1992), 'Alert and inert clients: The Scottish experience of parental choice of schools,' *Economics of Education Review* **11**(4), 339–350.

Winston, G. C. and D. J. Zimmerman (2003), 'Peer effects in higher education,' *NBER Working Paper* **9501**.

Wößmann, Ludger (2003), 'School resources, educational institutions and student performance: the international evidence,' *Oxford Bulletin of Economics and Statistics* **65**(2), 117–170.

Zimmer, R. W. and E. Toma (2000), 'Peer effects in private and public schools across countries,' *Journal of Policy Analysis and Management* **19**(1), 75–92.

Zimmerman, D. J. (2003), 'Peer effects in academic outcomes: Evidence from a natural experiment,' *Review of Economics and Statistics* **85**(1), 9–23.