Impacts of Within-Classroom Academic Heterogeneity and Ability-grouping on Teacher Practice

Authors:
Ha Yeon Kim, New York University hayeon@nyu.edu
Stephanie Jones, Harvard Graduate School of Education

Background

The stakes for effective classroom teaching are higher than ever. At present, in many states teachers' classroom instruction is frequently observed and evaluated, and teachers are held accountable for their students' achievement gains. One of the primary challenges in assessing quality of classroom teaching practice is that it likely varies a great deal according to the level and heterogeneity of the academic competence of the students (Guarino et al., 2014). Some teachers are assigned to teach a high- or low-achieving, relatively homogeneous group of students; others teach a group of students with widely varying academic competence.

Since Jeannie Oakes's seminal work (1985), much research has shown how schools exacerbate inequality by sorting academically homogeneous groups of students into different "tracks" (Bett, 2011; Schmidt et al., 2015; Slavin, 1993). Though this body of research raised awareness about the potentially detrimental effects of tracking among policy-makers and educators (Loveless, 1999), the practice has continued and remains widely prevalent. However, few studies have directly assessed its effect on teaching practice, nor used methods that permit causal interpretations of the impacts of ability-grouping on student outcomes.

Objectives

We capitalize a large dataset on teaching effectiveness and aim to test the impacts of within-classroom academic heterogeneity on teacher practice. The specific research aim is to test the impact of ability-grouping (*heterogeneous* versus *ability-grouped* classrooms) on observed teaching practice, adjusting for prior years' observed teacher practice.

Setting and Participants

The MET study sample included more than 3,000 teachers and their students (N = 44,500) from six mainly urban school districts across the country. Among them, 1,181 English and Mathematics teachers and their students (Grade 4-8) participated in the RCT study. This study includes teachers and students from fourth ($n = 146$) and fifth grade ($n = 156$) self-contained classrooms taught by generalist teachers (total N=302) participated in the Year 2 randomized proportion of the study.

Intervention

The Measures of Effective Teaching (MET) study is the largest study of instructional practice designed to validate and test measures of teaching effectiveness. In the 2009-10 non-experimental study year (Year 1), MET collected various measures of teaching quality from

teachers and students. These measures were then used to determine the *teaching effectiveness* of the teachers, the "treatment" that was randomly assigned to classrooms in the 2010-11 RCT year (Year 2). This study takes advantage of the random assignment of teachers to classrooms in the Year 2 of the MET study, and test the role of various level of within-classroom academic heterogeneity—when randomly assigned to teachers—on teaching practice, in absence of schools' strategic classroom-teacher matching practice.

Research Design

This study utilizes prior year's student state test scores to assess and generate the treatment contrast based on the level of **classroom heterogeneity**. Using the baseline student test scores, the *SD* of all students in each of the Year 2 classrooms are calculated. This score is used to create classroom heterogeneity scores. To estimate the average treatment effect (ATE) of classroom heterogeneity on observed teaching, an inverse-probability weighted regression adjustment technique is used. To do so, first, treatment and control groups is be defined. Considering the sample distribution of the classroom heterogeneity at baseline, this study determines a cut score. For example, the classrooms with a heterogeneity score higher than the mean will fall into the *treatment group* (=1, *heterogeneous*) and the rest of classrooms into the *control group* (=0, *ability-grouped*). Dichotomizing heterogeneity scores this way allows the identification of a clear counterfactual state, and an estimate simple to interpret, relative to the counterfactual (Caliendo & Kopeinig, 2008).

Data Collection and Analysis

For **classroom teaching practice,** we use the two measures of general classroom instruction quality: CLASS (Pianta et al., 2010) and the FFT (Danielson, 2011) (see Table 1 for domains and scores). Observation scores were coded from the video recordings of the participating teachers' classroom practice (> 4 recordings each year) in Year 1 and 2 and averaged across the recordings each year. In addition, various **covariates**, including student and teacher demographic variables and student test scores from the prior year, as well as classroom- and school-level covariates such as grade level, class size, school district, and classroom and school aggregates of student demographic characteristics and test scores will be included to adjust for potential selection bias.

Analyses will take place in two steps. First, the propensity scores (PS) for being assigned to the *heterogeneous* classrooms will be calculated by regressing the treatment variable on covariates in logit models. These models will include relevant teacher, classroom, and school characteristics as predictors, such as prior year teacher practice, class size, baseline aggregate test scores.

Then, these PS will be used to calculate inverse probability of treatment weight using the probability of being assigned to the treatment (=1/PS) and control group (1/(1-PS)). This approach allows weighting of each group to look similar to the full sample. Using these weights, OLS regression models with covariates will be run to estimate the ATE of heterogeneous classrooms on teaching practice.

Results

Preliminary analysis suggests that, controlling for other classroom and student composition characteristics including grade level, class size, % male, % ELL, % Hispanic and % black, within-classroom academic heterogeneity in ELA subject is positively related to teachers' emotional support and classroom management captured by CLASS (Table 2). Among the FFT scores, within-classroom academic heterogeneity was associated with a dimension measuring teacher practice of creating a classroom environment of respect and rapport (Table 3). Further analysis with imputed dataset and propensity score weighted regression approach will allow more accurate estimation of the role of classroom academic heterogeneity in teachers' classroom practice.

Conclusions

This study is one the few to examine the role of within-classroom academic heterogeneity and teacher practices in relation to student achievement. The findings from this study will inform teacher evaluation policies and educational practices to improve the accuracy and fairness of teacher evaluation and provide equal opportunities to learn for students of diverse academic competence.

Reference

Betts, J. R. (2011). The economics of tracking in education. In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (pp. 341–381). Elsevier.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*, 31–72.

Danielson, C. (2011). *Enhancing professional practice: A framework for teaching*. ASCD.

Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2014). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, *10*, 117–156.

Loveless, T. (2011). *The tracking wars: State reform meets school policy*. Washington, DC: Brookings.

Oakes, J. (1985). *Keeping track*. New Haven, CT: Yale University Press.

Pianta, R. C., Hamre, B. K., & Mintz, S. (2010). *Classroom Assessment Scoring System (CLASS) Manual: Upper Elementary*. Teachstone.

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The Role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher*, 371–386. doi:10.3102/0013189X15603982

Slavin, R. E. (1993). Ability grouping in the middle grades: Achievement effects and alternatives. *The Elementary School Journal*, *93*, 535–552. doi:10.1086/461739

Table 1

CLASS: Classroom Assessment Scoring System (Pianta et al., 2010)		
Emotional Support Domain	Classroom Organization Domain	Instructional Support Domain
Positive Climate	Behavioral Management	Instructional Learning Format
Teacher Sensitivity	Productivity	Content Understanding
Regard for Student Perspective	Negative Climate	Analysis and Problem Solving
		Quality of Feedback
		Instructional Dialogue
FFT: Framework for Teaching (Danielson, 2011)[a]		
Domain 2: Classroom Environment		Domain 3: Instruction
Creating an Environment of Respect & Rapport		Communicating With Students
Establishing a Culture for Learning		Using Questioning and Discussion Techniques
Managing Classroom Procedures		Engaging Students in Learning
Managing Student Behavior		Using Assessment in Instruction

Table 2
Regression models testing the role of classroom ELA academic heterogeneity in CLASS domain scores

	Emotional Support	Classroom Management	Instructional Support
	b (SE)	b (SE)	b (SE)
Classroom ELA Academic Heterogeneity (Prior Year SD)	0.24 (0.12) *	0.23 (0.12) *	-0.03 (0.13)
Classroom ELA Academic Average (Prior Year M)	0.00 (0.06)	0.06 (0.06)	-0.04 (0.06)
Grade Level (4th vs. 5th)	0.02 (0.04)	0.07 (0.03) *	0.00 (0.04)
Class size	0.01 (0.00) **	0.01 (0.00) *	0.01 (0.00) ***
% Male	-0.59 (0.23) *	-0.52 (0.22) *	-0.57 (0.25) *
% ELL	0.35 (0.15) *	0.12 (0.14)	0.32 (0.16) *
% Hispanic	-0.21 (0.11) *	-0.10 (0.11)	-0.27 (0.12) *
% Black	-0.26 (0.08) ***	-0.14 (0.08)	-0.38 (0.09) ***
Intercept	4.45 (0.20) ***	5.25 (0.20) ***	3.81 (0.22) ***
R^2	0.16	0.12	0.18
$F_{(8, 292)}$	7.05***	4.74***	8.23***

Table 3
Regression models testing the role of classroom ELA academic heterogeneity in FFT dimension scores

	Respect		Culture		Managing Procedure		Managing Behaviors		Communicate		Question		Engage		Assess	
	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)
Classroom ELA Academic Heterogeneity (Prior Year SD)	0.33	(0.11) **	0.05	(0.10)	0.13	(0.09)	0.14	(0.10)	0.07	(0.09)	0.02	(0.10)	0.03	(0.11)	-0.01	(0.10)
Classroom ELA Academic Average (Prior Year M)	0.05	(0.05)	0.02	(0.05)	0.06	(0.04)	0.07	(0.05)	0.07	(0.04)	0.02	(0.05)	0.00	(0.05)	-0.01	(0.05)
Grade Level (4th vs. 5th)	0.02	(0.03)	-0.03	(0.03)	0.07	(0.03) **	0.07	(0.03) *	-0.01	(0.03)	-0.02	(0.03)	-0.05	(0.03)	-0.04	(0.03)
Class size	-0.22	(0.21)	-0.17	(0.20)	-0.02	(0.18)	-0.35	(0.20)	-0.24	(0.18)	-0.06	(0.20)	-0.10	(0.21)	-0.12	(0.19)
% Male	0.19	(0.19)	0.27	(0.19)	0.13	(0.17)	0.13	(0.18)	0.06	(0.17)	0.03	(0.19)	-0.06	(0.20)	0.05	(0.18)
% ELL	0.16	(0.13)	0.11	(0.12)	0.05	(0.11)	0.09	(0.12)	0.06	(0.11)	0.13	(0.13)	0.17	(0.13)	0.05	(0.12)
% Hispanic	-0.22	(0.09) *	-0.40	(0.09) ***	-0.10	(0.08)	-0.17	(0.09)	-0.14	(0.08)	-0.26	(0.09) **	-0.40	(0.10) ***	-0.25	(0.09) **
% Black	-0.22	(0.07) **	-0.29	(0.07) ***	-0.11	(0.06)	-0.15	(0.07) *	-0.19	(0.06) ***	-0.29	(0.07) ***	-0.28	(0.07) ***	-0.27	(0.07) ***
Intercept	2.69	(0.16) ***	2.92	(0.15) ***	2.62	(0.14)	2.87	(0.15) ***	2.88	(0.14) ***	2.53	(0.16) ***	2.86	(0.17) ***	2.69	(0.15) ***
R^2	0.1		0.11		0.07		0.1		0.1		0.1		0.1		0.1	
$F_{(8, 292)}$	4.29		4.59		2.83		3.91		4.21		4.2		3.8		3.6	