

An Analysis of Accountability: Impact on Student Achievement

By Clara Peretz

Faculty Sponsor: Daniel Long, Wesleyan Sociology Department



QAC Summer 2012

Introduction

There has been growing skepticism surrounding the effectiveness of standardized tests and other assessments that use quantitative measures to draw conclusions about the quality of teaching that is taking place in classrooms across the United States (Gee, Haertel, Moss, Pullin, Young 2008, Elliot and Hout 2011). In recent years, with the No Child Left Behind Act of 2001, federal funds have been tied to standards-based assessments. However, many argue that it is difficult to address educator’s problems with a single type of evidence. Debate regarding what modes of school and teacher evaluation can best foster learning has highlighted concern about the loss of contextual information involved with using standardized tests to make cross-unit comparisons. Is the price of this loss of context worth the ability to make such comparisons?

In addition to understanding the specific relationship between learners and their learning environments, it is also important to recognize that learning occurs as part of a system that includes home, peer, and other influences (Gee, Haertel, Moss, Pullin, Young 2008). In light of the increasing emphasis placed on testing and the drastic consequences for teachers and schools failing to meet state-administered standards, we are interested in examining how differing accountability and incentive policies in U.S. high schools are linked with individual student achievement.

Data and Variables

The data in this analysis comes from the Educational Longitudinal Study of 2002, which traces a national sample of students beginning in tenth grade and follows their progression beyond high school. It is a multilevel study, with questionnaires administered to students, parents, math and English teachers, school principals, and heads of the school library media center. Schools are first selected and then tenth-grade students are randomly selected within each school.

The analyses look at the relationships between student achievement in math and measurements of teacher and principal accountability, as the main independent variables of interest. The simple models include just these variables that are concerned with accountability, analyzed with two different dependent variables, math IRT gain and math IRT estimated number right, based on students’ 10th grade scores. The independent variables of interest include modes of teacher evaluation, whether teachers are evaluated by the principal or administrators, by other teachers, or by students. Incentives, in the form of recognition given to good teachers, are also included in the models. These include special awards, better students, a lighter teaching load, relief of duties, priority on requested materials, and higher pay. The final grouping of variables in this model pertains to the principal’s beliefs about the influence of the following on how his or her performance is evaluated by superiors: student test scores, school environment, efficiency of administration, parent involvement, relationship with the community, and new programs and reform.

Covariates were grouped into blocks based on source of influence. The following groups of variables were created: demographic characteristics, family characteristics, friends’ influence, teacher qualifications & beliefs, and student characteristics. Some constructs were created by combining responses to similar questions or sub-questions from a multi-part question.

Methods

Initially we conducted a series of OLS regressions to look at the association between measures of accountability and math achievement. However, there were a significant number of missing observations in our data set. To account for missing data, we performed multiple imputations to increase the sample size. Using this larger data set, we created multi-level models, nesting students within schools.

Results

Grade 10 IRT Math Scores					
	Accountability with no covariates (1)	Accountability with student-level SES (2)	Accountability with school and student-level SES (3)	Accountability with school and student-level SES and demographic controls (4)	Accountability with school and student-level SES and all controls (5)
<i>Teacher</i>					
<i>Accountability</i>					
Evaluations by Principal	0.92 [1.29]	1.12 [1.16]	1.50 [1.04]	1.91* [0.96]	1.99* [0.83]
Evaluations by Teacher	1.27** [0.45]	0.83* [0.39]	0.21 [0.36]	0.21 [0.34]	0.72* [0.30]
Evaluations by Student	2.46*** [0.63]	1.30* [0.51]	0.15 [0.47]	0.11 [0.41]	-0.11 [0.38]
<i>Teacher Incentives</i>					
Special Awards	-0.71 [0.37]	-0.71* [0.31]	-0.72** [0.28]	-0.51 [0.27]	-0.56* [0.24]
Better Students	0.26 [0.64]	0.45 [0.54]	0.44 [0.49]	0.37 [0.45]	-0.21 [0.39]
Lighter Teaching Load	-0.45 [1.73]	-1.18 [1.38]	-1.48 [1.17]	-1.39 [1.10]	-0.98 [0.82]
Duty Relief	1.65 [2.20]	1.71 [1.74]	1.75 [1.52]	1.78 [1.40]	1.15 [1.15]
Priority on Material Requests	-1.56 [1.16]	-1.38 [1.07]	-1.21 [0.96]	-0.98 [0.84]	-0.33 [0.66]
Higher Pay	-0.15 [0.89]	-0.56 [0.79]	-1.42* [0.72]	-1.06 [0.69]	-0.51 [0.46]
<i>Principal Accountability</i>					
Test Scores	-1.38*** [0.38]	-1.02** [0.32]	-0.40 [0.29]	-0.18 [0.28]	-0.42 [0.24]
School Environment	-0.54 [0.65]	-0.24 [0.53]	-0.06 [0.48]	-0.16 [0.45]	-0.19 [0.37]
Efficiency of Administration	1.59* [0.79]	1.32* [0.65]	1.07 [0.57]	1.09* [0.51]	0.62 [0.46]
Parent Involvement	1.08* [0.51]	0.59 [0.39]	0.17 [0.35]	0.33 [0.33]	-0.0006 [0.26]
Relationship with Community	0.07 [0.61]	0.04 [0.52]	0.11 [0.47]	-0.26 [0.44]	-0.22 [0.37]
New Programs/ Reform	-1.19* [0.54]	-1.02* [0.41]	-0.85* [0.37]	-0.58 [0.32]	-0.22 [0.28]
Student-level SES		5.50*** [0.13]	4.95*** [0.13]	0.73*** [0.09]	2.04*** [0.14]
Mean school SES			6.07*** [0.39]	1.21*** [0.188]	2.80*** [0.37]
Constant	36.88*** [1.46]	36.61*** [1.30]	36.11*** [1.17]	5.56*** [0.63]	28.90*** [1.96]
BIC	123606.666	121819.706	121598.114	100560.97	115539.85
N	16197	16197	16197	16197	16197

IRT Math Gain Scores					
<i>Teacher</i>					
<i>Accountability</i>					
Evaluations by Principal	-0.64 [0.61]	-0.56 [0.56]	-0.47 [0.55]	-0.19 [0.54]	0.21 [0.63]
Evaluations by Teacher	0.50* [0.20]	0.38* [0.19]	0.27 [0.19]	0.29 [0.19]	0.14 [0.20]
Evaluations by Student	0.63* [0.27]	0.37 [0.26]	0.13 [0.26]	-0.01 [0.27]	-0.07 [0.29]
<i>Teacher Incentives</i>					
Special Awards	-0.18 [0.14]	-0.20 [0.13]	-0.21 [0.13]	-0.21 [0.13]	-0.13 [0.13]
Better Students	-0.08 [0.29]	-0.04 [0.28]	-0.04 [0.27]	-0.03 [0.27]	-0.20 [0.28]
Lighter Teaching Load	0.36 [0.58]	0.27 [0.54]	0.23 [0.53]	-0.07 [0.54]	0.25 [0.50]
Duty Relief	0.75 [0.59]	0.79 [0.57]	0.82 [0.56]	0.92 [0.55]	0.70 [0.55]
Priority on Material Requests	-0.39 [0.45]	-0.33 [0.44]	-0.28 [0.43]	-0.21 [0.43]	-0.02 [0.41]
Higher Pay	-0.07 [0.42]	-0.25 [0.40]	-0.44 [0.40]	-0.50 [0.39]	-0.60 [0.36]
<i>Principal Accountability</i>					
Test Scores	-0.31 [0.21]	-0.19 [0.20]	-0.07 [0.20]	-0.11 [0.20]	-0.01 [0.19]
School Environment	0.08 [0.30]	0.11 [0.28]	0.14 [0.28]	0.32 [0.28]	0.25 [0.27]
Efficiency of Administration	-0.31 [0.32]	-0.37 [0.31]	-0.41 [0.31]	-0.37 [0.30]	-0.39 [0.30]
Parent Involvement	0.94*** [0.22]	0.84*** [0.22]	0.76*** [0.22]	0.63** [0.22]	0.55* [0.22]
Relationship with Community	-0.35 [0.23]	[-0.35] [0.2]	-0.34 [0.22]	-0.24 [0.22]	-0.13 [0.23]
New Programs/ Reform	-0.11 [0.16]	-0.07 [0.15]	-0.02 [0.15]	-0.08 [0.15]	0.02 [0.15]
Student-level SES		0.96*** [0.08]	0.73*** [0.09]	0.73*** [0.09]	0.27* [0.11]
Mean School SES			1.141*** [0.18]	1.21*** [0.19]	0.23 [0.26]
Constant	5.87*** [0.62]	5.79*** [0.58]	5.68*** [0.58]	5.56*** [0.63]	0.439 [1.89]
BIC	106385.772	106229.632	106173.606	88867.3162	106264.522
N	16197	16197	16197	16197	16197
*p<0.05, **p<0.01, ***p<0.001					
Notes:					
–Model 4 controls for demographic characteristics					
–Model 5 controls for demographic, school, family, friends, teacher, and student characteristics					
–Standard error in brackets					
–The boxed model represents the best fit model					

Conclusions

In Model 1, using grade 10 IRT math scores as the dependent variable, the independent variables of interest, teacher accountability based on evaluations by other teachers and by students, and principal accountability influenced by administrative efficiency and by parent involvement with the school, appears to be positively associated with grade 10 math scores. We also find a negative association between the dependent variable and principal accountability influenced by students’ test scores and by principal’s reform policies. At first glance, this analysis provides an argument against the accountability policies of No Child Left Behind. However, upon further investigation, adding controls invalidates the initial findings.

After student-level socioeconomic status (SES), school-level SES, and demographics are controlled for in Model 4, teacher evaluations of teachers, as well as the influence of administrative efficiency on principal accountability, becomes significant again after disappearing in Model 3, and evaluations of teachers by principals becomes significant as well. This suggests that the positive effect of these variables is masked by the racial composition of a school. These modes of evaluation tend to occur in high-minority schools, and high-minority schools tend to have lower achievement in math. The negative effect of reform policies as an influence on principal accountability loses significance in the Model 4, suggesting that lower math gain scores are associated with high-minority schools, rather than with the pressure on principals to be enacting reform policies.

Although they provide notable results, the models explained above do not control for previous achievement. Analysis using a gain score variable revealed that evaluations of teachers by other teachers and by students, as well as parent involvement as an influence on principal accountability, have a positive effect on math gain, as shown in Model 1 when IRT math gain is used as the dependent variable.

When SES is added in Model 2, evaluations by students disappears. In Model 3, when school-level SES is controlled for, the significance of teacher evaluations by teachers disappears, which means that this mode of evaluation is most prevalent in high-SES schools and that the SES of the school is responsible for the students’ higher gain scores, rather than the teacher accountability policies practiced by the school.

In Model 4, when demographic controls are introduced, there is no change in the significance of the accountability variables. Parent involvement remains significantly, positively associated with math gain score across all the models.

In using both the grade 10 IRT score and the gain score as dependent variables, including all the covariates does not improve the overall fit of the model. Therefore, Model 4 is concluded to be our best fit model.

References

Gee, Haertel, Moss, Pullin, Young. (2008). *Assessment, Equity, and Opportunity to Learn*. New York, NY: Cambridge University Press.

National Research Council. (2011). *Incentives and Test-Based Accountability in Education*. Committee on Incentives and Test-Based Accountability in Public Education, M. Hout and S.W. Elliott, Editors. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Acknowledgements

I would like to thank my sponsoring professor, Daniel Long, and QAC director, Manolis Kaparakis, for providing me with this opportunity, and for their help throughout the process; I would have been lost without them. A special thanks to Tim Wolock for his enthusiastic support, attentiveness, and patience.