Foley, B. P. (2014). Modeling the relationships among measures of teacher quality and student performance in high school geometry. *The Researcher, 26*(1), 28-33.

Modeling the Relationships Among Measures of Teacher Quality and Student Performance in High School Geometry

Brett P. Foley

Alpine Testing Solutions, Orem, UT

Abstract: *This study evaluates the relationships among measures of teacher quality and student performance in high school geometry. The results indicate that student readiness is a much stronger predictor of student end of course scores than are teacher quality measures. Implications for teacher evaluation systems are discussed.*
Keywords: *Assessment; Geometry; Teacher Evaluation*

INTRODUCTION

The use of scores from standardized tests of student achievement in K-12 teacher evaluation systems is gaining support among U.S. policymakers and the public (e.g., Bushaw & Lopez, 2012; Race to the Top Fund, 2009). However, professionals in the testing community have expressed reservations at using student data as a component of teacher evaluations (Baker et al., 2010), and in some cases teacher evaluations utilizing student test scores have been struck down by the courts (Postal, 2012). The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), along with current test validity theorists (e.g., Kane, 2013) call for the collection of evidence to support uses or interpretations of test results. But how should such evidence be collected? It makes intuitive sense that if student achievement data is appropriate for measuring teacher performance, these data should be correlated with other measures of teacher quality. That is the purpose of this study was: to evaluate the relationships between several measures of teacher quality and student performance, focusing on the educational context of high school geometry. Research questions this study addresses include:

1. How do different measures of teacher quality correlate with student performance in geometry?
2. Can a statistical model be constructed that identifies a set of teacher quality measures that predicts student performance in geometry beyond what is explained by the students' previous geometry knowledge?

REVIEW OF THE LITERATURE

The appropriateness of using student to test scores for evaluating teachers has become a subject of debate among policymakers and the professional psychometric community. Similarly, researchers examining the relationships between measures of teacher quality and student test scores have found mixed results.

Correspondence concerning this article should be addressed to Brett P. Foley, 10345 SW 119th St., Denton, NE 68339. Email: brett.foley@alpinetesting.com

As previously mentioned, there is increasing pressure from policymakers and the public to use student test scores to evaluate teachers (e.g., Bushaw & Lopez, 2012; Race to the Top Fund, 2009). However, leaders in the psychometric community have advocated for limited use of student test results (especially as evinced by value-added models) in teacher and school evaluation systems (e.g., Baker et al., 2010; Darling-Hammond, 2012; Hartel, 2013; Linn, 2006). Empirically, some studies have found little to no relationship between measures of teacher quality (e.g., education, experience, certification status, National Board Certification) and student achievement (e.g., Muñoz & Chang, 2008). Other studies have shown inconclusive or mixed results (e.g., Goldhaber & Anthony, 2007; Hacke, 2010; Harris & Sass, 2008; Kane, Rockoff, & Staiger, 2006); still others have identified teacher characteristics that are positively related with student achievement (e.g., Gersten, Baker, Haager, & Graves, 2005; Wayne & Youngs, 2003). The lack of agreement between policymakers and psychometric leaders, as well as inconsistent results from empirical studies across different grades, subjects, and teacher quality measures, point to the need for additional evaluation of the relationships between measures of teacher quality and student achievement.

RESEARCH METHODOLOGY

This study is based on data from a sample of approximately 128 secondary mathematics teachers in the Appalachian Mathematics Science Partnership (AMSP), the East Alabama Mathematics and Science Partnership Alabama TEAMS project, and a high school in Las Vegas, NV. The outcome variable for this study is average student scores for each teacher on an end-of-course geometry assessment (Kentucky Department of Education, 2010; possible scores: 0-40) covering four major areas of geometry: similarity, congruence, area, and volume (hereafter, "student EOC scores"). The measures of teacher quality include the following:

- **Teacher Knowledge:** This is measured by the teachers' performances on the Geometry Assessment for Secondary Teachers (GAST; Davidson et al., 2012; possible scores: 0-45), which provides measures of teachers' content and pedagogical knowledge of geometry (hereafter, "knowledge").
- **Teacher Experience & Education:** This is be measured by teacher demographic variables (i.e., years of experience and attainment of advanced education degrees; hereafter, "experience" and "education," respectively).
- **Classroom Performance:** Teachers were observed in the classroom and their classroom behaviors recoded on an evaluation form. Their observations were then summarized into an index of the level of cognitive challenge of classroom instruction provided (Davidson et al., 2012; hereafter, "classroom performance"; possible scores 1-3).

Because prior knowledge plays an essential role in how well and how much students will learn (Bruning, Schraw, Norby, & Ronning, 2004), it is important to control for students' prior knowledge when examining outcomes. This is accomplished by including the average student score for each teacher on a measure of geometry readiness (Usiskin & Senk, 1982; possible scores: 0-20) as a covariate in the analyses (hereafter, "student readiness").

Data are analyzed using correlational and regression methods. Along with standard statistical tests for bivariate semi-partial correlations, teacher quality variables are be entered into a stepwise regression model (with student readiness) to identify the variables (if any) that best predict student performance.

RESULTS

Characteristics of the teachers participating in the study are included in Table 1. Approximately 79% of the teachers held advanced degrees and the average amount of teaching experience was approximately 13 years. Student readiness test results (for the average teacher) were collected from approximately 22 geometry students; on average, a similar number of students completed the EOC test (about 21 students per participating teacher). Because of missing data on some measures, the number of teachers with data on each characteristic range from 109-122; sample sizes for the correlational and regression analysis ranging from 108-121 (using listwise deletion).

Table 1. *Characteristics of Participating Teachers (N = 109-122)*

Characteristic	*M*	*SD*
Experience (years)	13.10	8.99
Knowledge (score, 0-45 scale)	24.09	7.41
Classroom Performance (score, 1-3 scale)	1.60	0.18
Student Readiness (score 0-20 scale)	14.54	2.60
Student EOC Scores (0-40 scale)	15.64	5.72
Number of students who took readiness test	22.39	8.75
Number of students who took EOC test	20.96	8.26
Education (% with advanced degrees)	78.51	

Most measures of teacher quality were positively correlated with each other (see Table 2). Teachers with advanced degrees tended to have more experience, higher geometry/pedagogy scores, and higher classroom performance ratings. Additionally, teachers with more experience tended to have higher geometry/pedagogy scores.

Table 2. *Correlations Among Teacher Quality Measures*

Measure	1	2	3	4
1. Education	_			
2. Experience	.29**	_		
3. Knowledge	.18*	.31**	_	
4. Classroom Performance	.26**	.06	.09	_

Note. Significance values are based on one-sided tests because of the a priori hypothesis that teacher quality measures would be positively correlated. $^{*}p < .05$. $^{**}p < .01$.

When analyzed in isolation, all teacher quality variables except for classroom performance were significantly positively correlated with student EOC scores (see Table 3). However, the variable most strongly correlated with student EOC scores was student readiness ($r = .84$, $p < .001$). When student readiness was accounted for, the correlations between teacher quality measures and student EOC scores dropped substantially, with only experience and knowledge remaining significant. Taken together, the four measures of teacher quality accounted for 22% of the variability in student EOC scores (ignoring student readiness; R^2=.22, R^2_{adj}=.19). When student readiness was included in the model, 74% of the variability in student EOC scores was accounted for (R^2=.74, R^2_{adj}=.73). When accounting for the other measures of teacher quality and student readiness, only teacher experience remained a significant predictor of student EOC scores. As a group, the measures of teacher quality only explain about 4% additional variance in student EOC scores after accounting for student readiness; student readiness was a much stronger predictor of student EOC scores than teacher quality measures.

Table 3. *Bivariate and Semipartial Correlations Among Teacher Quality Measures and Student EOC Scores*

		Measure(s) accounted for		
Measure	Bivariate correlations	Student Readiness[1]	All other measures[1] *excluding Student Readiness*	All other measures[1]
1. Education	.21*	.01	.02	-.06
2. Experience	.40**	.15**	.35**	.15**
3. Knowledge	.23**	.11*	.09	.07
4. Classroom Performance	.16	.09	.14	.10*
5. Student Readiness	.84**	-	-	.72**
Variance collectively explained by set of predictors			R^2 = .22** R^2_{adj}=.19	R^2 = .74** R^2_{adj}=.73

Note. Significance values are based on one-sided tests because of the a priori hypothesis that teacher quality measures would be positively correlated with Student EOC Scores
[1]Semipartial correlations. **p* < .05. ***p* < .01.

The teacher quality measures and student readiness were entered into a stepwise regression model to try and identify a parsimonious set of predictors of student EOC scores. The resulting model contained only student readiness and teacher experience as predictors and explained 72% of the variability in student EOC scores (see Tables 4 and 5; R^2=.72, R^2_{adj}=.72).

Table 4. *Stepwise Regression Results for Full Set of Predictive Measures of Student EOC Scores*

Step	Measure added to model	Total R^2	Total R^2_{adj}	Incremental R^2
1	Student Readiness	.695	.692	.695
2	Experience	.722	.717	.026

Note: All values significant at .01 level. Threshold for predictors entering the model = .05. Threshold for predictors being removed from the model = .10. Predictor variables included in the analysis: Education, Experience, Knowledge, Classroom Performance, Student Readiness.

Table 5. *Stepwise Regression Analysis Summary for Predictive Measures of Student EOC Scores*

Measure	*B*	*SEB*	*β*	*t*	*p*
Intercept	-10.54	1.66		-6.35	<.001
Student Readiness	1.70	0.12	.78	14.28	<.001
Experience	0.10	0.03	.17	3.16	.002

Importance to Field

As policymakers push for additional accountability for the teaching profession, more pressure is brought to bear on teacher evaluation systems. There is growing support for using student achievement as a component of these systems. However, to help ensure the validity of such interpretations, it is important to evaluate the relationship between student data and other characteristics of effective teachers. As such, this research is timely and essential in that it serves the important function of evaluating how measures of teacher quality relate to student performance. The weak relationships found between the teacher quality measures and student EOC scores support the recommendations of the psychometric community that if student test scores are going to be used to evaluate teachers, they should only be used as part of a more comprehensive evaluation system.

The results of this study are limited in their generalizability in that only a single subject area was evaluated and the sample size was relatively small. Additionally, because the assessments used in this study had no stakes attached for students or teachers, the level of examinee motivation may not be the same as for more high-stakes assessments. Because this research is part of an ongoing research program, the analyses performed were rudimentary and the results should be considered preliminary. More sophisticated analyses such as hierarchical linear modeling and structural equation modeling would likely produce more precise results in that these methodologies account for classroom level variability and the allow for correction of the attenuation caused by measurement error.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Baker, E. L., et al. (2010). *Problems with the use of student test scores to evaluate teachers* (Briefing Paper No. 278). Washington, DC: Economic Policy Institute.

Bruning, R. H., Schraw., G. J., Norby, M. N., & Ronning, R. R. (2004). *Cognitive psychology and instruction* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Bushaw, W. J., & Lopez, S. J. (2012). *Public education in the United States: A nation divided.* Bloomington, IN: Phi Delta Kappan.

Darling-Hammond, L. (2012). Value-added teacher evaluation. *Education Week, 31*(24), 24, 32.

Davidson, A. H, Buckendahl, C. W., Stokes, M., O'Leary, L., Goodwin, C., & Foley, B. P. (2012). *Geometry Assessment for Secondary Teachers project technical manual*. Unpublished manuscript, Alpine Testing Solutions, Orem, UT.

Gersten, R., Baker, S. K., Haager, D, & Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners: An observational study. *Remedial and Special Education, 26*(4), 197-206.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*, 134-150.

Hacke, W. (2010). *Meta-analysis comparing student outcomes for National Board Certified teachers and non-National Board Certified Teachers.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3424858)

Harris, D. N., & Sass, T. R. (2008). *Teacher training, teacher quality, and student achievement* (Working paper 3). Washington, DC: National Center for Analysis of Longitudinal Data in Educational Research.

Hartel, E. H. (2013). *Problems with the use of student test scores to evaluate teachers* (William H. Angoff Memorial Lecture Series #14). Princeton, NJ: Educational Testing Service.

Linn, R. L. (2006). Validity of inferences from test based educational accountability systems. *Journal of Personnel Evaluation in Education, 19,* 5-15.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1-73.

Kane, T. J., Rockoff, J. E., Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City* (Working paper 12155). Cambridge, MA: National Bureau of Economic Research.

Kentucky Department of Education. (2010). *End-of course assessments in high school mathematics*. Frankfort, KY: Author.

Muñoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education, 20,* 147-164.

Postal, L. (2012, August 22). Judge strikes down state merit pay rules as 'invalid.' *Orlando Sentinel*. Retrieved from http://www.orlandosentinel.com

Race to the Top Fund, 74 Fed. Reg. 59688 (2009) (to be codified at 34 CFR Subtitle B, Chapter II).

Usiskin, Z., & Senk, S. (1982). *Van Hiele levels and achievement in secondary school mathematics*. Chicago: The University of Chicago.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89-122.