The Relationship between Teacher Mathematics Knowledge and Teacher Practice

Mark C. Greenwood[1], Elizabeth Burroughs[1], David Yopp[1], Megan Higgs[1], and John Sutton[2].

Abstract

Relationships between subject area knowledge of teachers and teacher practice are of interest in understanding the preparation and professional development of teachers as well as in assessing teacher performance. The study explores the relationship between Mathematics Knowledge for Teaching (MKT) and Inside the Classroom – Classroom Observation Protocol (ITC-COP) in a sample of 129 teachers from 25 school districts across seven states. Using linear and ordinal mixed models to account for a random district effect on ITC-COP measurements, modest relationships were identified between MKT and seven of eight different aspects of the ITC-COP.

Introduction

This study investigates claims that teachers' mathematics knowledge for teaching explains facets of teacher practice. It uses a sample of 129 teachers from 25 school districts across seven states to explore relationships between data from a commonly used, valid and reliable multiple-choice measure of teacher mathematics knowledge for teaching (the Mathematics Knowledge for Teaching, MKT, Hill et al., 2004) and data from a commonly used, valid and reliable classroom observation protocol (the Inside the Classroom – Classroom Observation Protocol, ITC-COP, Horizon, Inc., 2000). The analysis explores the question whether data about teacher mathematics knowledge is useful in explaining elements of teacher practice and compares teaching practice ratings between low, medium, and high knowledge teachers.

This research situates well among recent studies (Hill et al., 2008; Hill et al., under review; and Stein and Kaufman, 2010) that have also investigated the relationship between teacher knowledge and practice. This research differs from these recent studies because it uses a larger sample of teachers, a more diverse setting (in terms of district type and curriculum used), and two instruments that are widely used but not necessarily designed to work in concert.

While it stands to reason that teacher mathematics knowledge influences mathematics teaching practice, three decades of research have failed to explain the nature and strength of these relationships. Recent studies (e.g. Hill et al., 2008; Hill et al., under review; and Stein and Kaufman, 2010) have contributed greatly to this understanding by using more objective measures of teacher mathematics knowledge that align well with teachers' actual practice and sample sizes larger than many earlier studies (from 10 to 48 participants and multiple classroom observations for each participant). But, the question remains open regarding to what degree a teacher's score on a mathematics content assessment can predict teacher practice, when measures are instruments commonly used to assess professional development and not necessarily designed

[1] Department of Mathematical Sciences, Montana State University, Bozeman, MT
[2] RMC Research Corporation, Denver, CO

to work in concert for a given study. The present study contributes to the growing understanding of how teacher knowledge relates to teacher practice by addressing the following:

1. Can a paper and pencil assessment of a teacher's mathematics knowledge for teaching significantly explain aspects of the teacher's classroom practice, as measured by a classroom observation protocol?
2. If teachers are binned as displaying low, medium, or high levels of mathematics knowledge for teaching, are there significant differences in the practices of teachers in these bins?

Methods

This study uses data from n=129 teachers collected during 2010 as part of a larger study on elementary and middle school mathematics teachers mathematics knowledge and teaching practice. Twelve observers engaged in reliability training to standardize the ITC-COP observations. The observers were then allocated to observe the 129 teachers in the study based on geographical efficiency, which tended to correlate them highly but not perfectly with the 25 districts that had teachers participating in the study. This is not a nationally representative random sample but is diverse in both geography and teacher experience as participation in the study was based on other factors. The ITC-COP measurements involved announced classroom visits for lessons related to number sense or algebraic reasoning that took place between March and May 2010. Between January and March 2010, the teachers completed the Mathematics Knowledge for Teacher Survey (MKT), providing an item response theory latent ability score. It was selected because of its validity and reliability (Hill et al., 2004) and its alignment with content K-8 teachers actually teach (Ball et al, 2008). Hill et al. (2008) discuss the merits of using the MKT for this type of study. Teachers took one of two randomly selected versions of the MKT at their level of instruction. The scores on the two versions were equated using the original data sets used to develop the instruments.

The ITC-COP (Horizon Research, Inc., 2000) contains 29 frequency based ratings across four labeled subject areas (see Table 1 below for the questions and organization of the questions), with an overall synthesis rating for each area and an overall capsule rating for the entire session. The 29 frequency based ratings were subjected to an exploratory factor analysis using maximum likelihood estimation with varimax rotation. A combination of scree plots and exploration of eigenvalues was used to select the number of factors and estimate the factor analysis solution. Three underlying factors were identified that related to teaching skills, content knowledge, and collaboration. Ten items were removed because of problematic factor structure coefficients. The three factors had Cronbach's Alpha between 0.875 and 0.939. For the identified factors, factor scores based on the average of responses on questions identified as loading highly on each factor were calculated. The synthesis and capsule ratings were all treated as ordinal variables and begin as measurements on a five-level Likert scale. The capsule rating is then additionally refined for ratings of “3” to “Low 3”, “3”, or “High 3”. These responses were converted into a seven-level Likert scale to respect this ranking of the intermediate responses.

In the ITC-COP responses, there is a potential for district level impacts on teacher practice due to differences in curriculum, training, and demographics of teachers so all analyses account for the effect of districts to conditionally evaluate the impact of MKT on ITC-COP. This effect could also be attributed to an observer effect based on the design of the study despite efforts to control inter-rater variability because the observers and districts are confounded. Eight different responses are examined for the ability of MKT to explain responses on different aspects of the ITC-COP, controlling for random district effects, using mixed model techniques (Pinheiro and Bates, 2000; Agresti, 2010). The factor scores are modeled using linear mixed models in R (R Development Core Team, 2011) based on the lme4 package (Bates et al., 2011). For the ordinal synthesis and capsule ratings, ordered probit mixed models are used from the ordinal2 package (Christensen, 2011). Probit link models were used to enhance numerical stability of the models and to provide a more natural interpretation of the effects, which were assumed to be constant across response categories.

Hill et al. (under review) suggest that the extremes of MKT may impact teacher practice differently than the results that are observed for more average MKT scores. However, their analysis assumes a linear relationship between MKT and their measure of teacher practice. To directly assess the difference between levels of MKT, it was binned into low, medium, and high categories based on the assumed underlying normal distributions of latent abilities, placing theoretically 25 percent into the lowest and highest groups and the middle 50 percent into a middle level group. The observed MKT scores for this sample of teachers along with these cut-offs are displayed in Figure 1. Follow-up tests to compare the three contrasts between different pairs of MKT levels were conducted with a Bonferroni adjustment to control for accumulation of error across the three tests within each model.

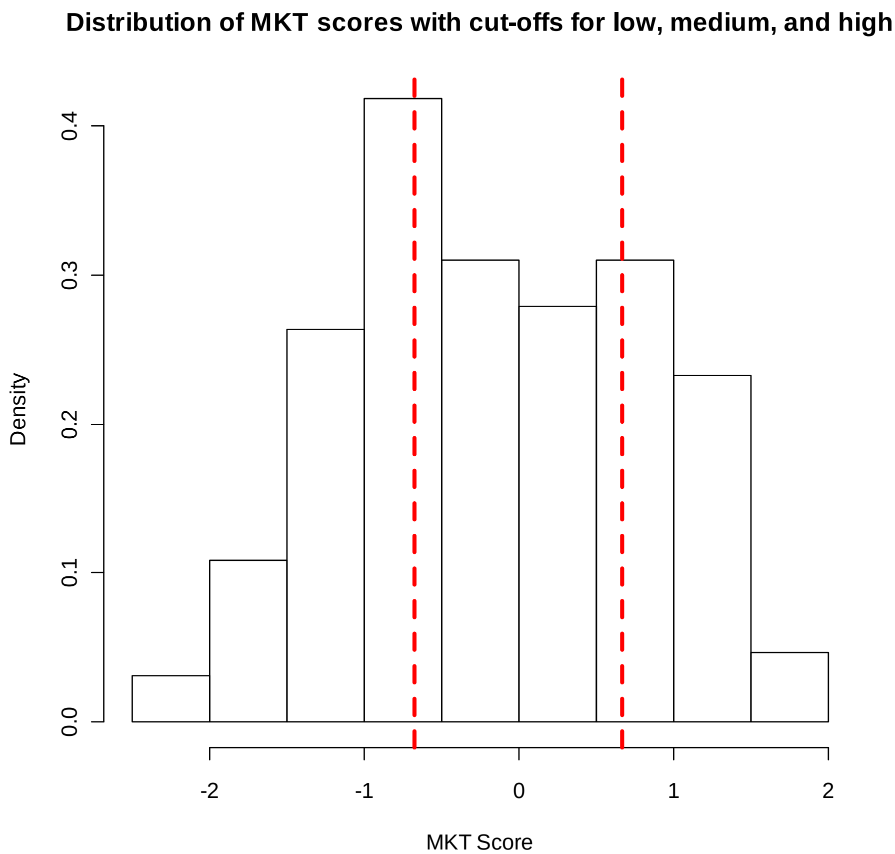

Figure 1. Histogram of MKT scores with cut-offs to bin the assumed underlying normal distribution into the lowest 25%, middle 50%, and highest 25% groups.

The linear mixed model for each of the three factor scores for teacher *i* in district *k* is $score_{i(k)} = \beta_0 + \beta_1 x_{Mi(k)} + \beta_2 x_{Hi(k)} + u_k + \varepsilon_{i(k)}$. β_0 is the value for the low MKT group, β_1, the deviation between low and medium MKT groups with $x_{Mi(k)} = 1$ for a medium MKT, and β_2, the deviation between low and high MKT groups with $x_{Mi(k)} = 1$ for a high MKT. The random effect for district k is assumed to be $u_k \sim N(0, \sigma_u)$ and the residual random teacher level error is $\varepsilon_{i(k)} \sim N(0, \sigma_\varepsilon)$. The ordinal mixed model for the j^{th} response category for the i^{th} teacher in the k^{th} district is $\text{probit}[P(y_{i(k)} \leq j)] = \alpha_j - u_k - \beta_1 x_{Mi(k)} - \beta_2 x_{Hi(k)}$ where $u_k \sim N(0, \sigma_u)$ is the random district effect, α_j are the J-1 thresholds between the categories, and $x_{Mi(k)}$ and $x_{Hi(k)}$ are defined as before.

Results

The factor analyses of the 29 frequency-based measurements of ITC-COP are described in Table 1. The first factor is interpreted as mathematics content knowledge, the second factor as student centeredness, and the third factor as student collaboration. The presence of only three identified factors contrasts with the four groups of questions labeled as Design, Implementation, Content, and Classroom, which also produce four synthesis responses for each labeled group of questions that are analyzed separately below.

Table 1. Factor Pattern/Structure Matrix based on maximum likelihood estimation and varimax rotation of the ITC-COP frequency based rating items.

Item	Factor 1: Mathematics Content Knowledge	Factor 2: Student Centeredness	Factor 3: Student Collaboration	h^2
D1. The design of the lesson incorporated tasks, roles, and interactions consistent with investigative math.	0.57		0.59	0.76
D2. The design of the lesson reflected careful planning and organization.	**0.65**	0.36	0.35	0.68
D3. The instructional strategies and activities used in this lesson reflected attention to students' experience, preparedness, prior knowledge, and/or learning styles.	**0.56**	0.45	0.41	0.68
D4. The resources available in this lesson contributed to accomplishing the purposes of the instruction.	**0.64**		0.37	0.62
D5. The instructional strategies and activities reflected attention to issues of access, equity, and diversity for students.		0.38	**0.63**	0.63
D6. The design of the lesson encouraged a collaborative approach to learning among students.			**0.82**	0.79
D7. Adequate time and structure were provided for "sense-making."	**0.71**	0.39		0.74
D8. Adequate time and structure were provided for wrap-up.	0.5	0.33		0.41
I1. The instructional strategies were consistent with investigative math.	0.55	0.44	0.51	0.75
I2. The teacher appeared confident in his/her ability to teach math.	**0.58**	0.35		0.49
I3. The teacher's classroom management style/strategies enhanced the quality of the lesson.	0.39	**0.7**		0.73
I4. The pace of the lesson was appropriate for the developmental needs of the students and the purposes of the lesson.	0.54	0.5		0.62
I5. The teacher was able to "read" the students' level of understanding and adjusted instruction accordingly.	0.51	0.61		0.7
I6. The teacher's questioning strategies were likely to enhance the development of student conceptual understanding/problem-solving.	0.56	0.55	0.35	0.73
C1. The math content was significant and worthwhile.	**0.63**		0.37	0.58
C2. The math content was appropriate for the developmental levels of the students in this class.	**0.51**	0.35	0.39	0.54
C3. Teacher-provided content information was accurate.	**0.61**	0.41		0.55
C4. Students were intellectually engaged with important ideas relevant to the focus of the lesson.	0.53	0.52	0.46	0.76
C5. The teacher displayed an understanding of math concepts.	**0.69**	0.41		0.67

C6. Math was portrayed as a dynamic body of knowledge continually enriched by conjecture, investigation analysis, and/or proof/justification.	0.49	0.48	0.31	0.57
C7. Elements of math abstraction were included when it was important to do so.	**0.68**			0.55
C8. Appropriate connections were made to other areas of math, to other disciplines, and/or to real-world concepts.			0.35	0.28
C9. The degree of "sense-making" of math content within this lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson.	0.64	0.5		0.75
CC1. Active participation of all was encouraged and valued.		**0.65**	0.48	0.71
CC2. There was a climate of respect for students' ideas, questions, and contributions.	0.44	**0.67**	0.33	0.76
CC3. Interactions reflected collegial working relationships among students.			**0.81**	0.74
CC4. Interactions reflected collaborative working relationships between teacher and students.	0.45	**0.64**	0.32	0.72
CC5. The climate of the lesson encouraged students to generate ideas, questions, conjectures, and/or propositions.	0.44	**0.65**	0.33	0.72
CC6. Intellectual rigor, constructive criticism, and the challenging of ideas were evident.	0.43	**0.61**	0.41	0.73
Eigenvalues	17.42	1.52	1.02	
% of variance after rotation	27.20%	21.30%	16.80%	

Note: h^2 = communality coefficient. Loadings in bold mean that that item is attributed to that factor.

Results of the mixed model analyses are presented in Table 2, starting with ordinal response mixed models for the first five responses and linear mixed models for the three factor scores. For the ordinal mixed models, the MKT effects can be interpreted as shifts in standard deviations of the underlying latent trait (manifested in the ordinal scale). For the linear mixed models, the response scale is the factor score that ranged between one and five. The capsule rating is a unique observation that attempts to combine aspects of all the other measurements. For the capsule rating, the estimated difference between high and low MKT groups is 0.63, suggesting that the high MKT group scores 0.63 standard deviations higher on the underlying continuous trait than the low group. Between medium and low, the difference is only 0.31. In general, this is not a very large magnitude of an effect and leads to suspicion about the practical significance of the effects, regardless of their statistical significance. Similarly effect sizes were found for the remainder of the ordinal-response models, always with higher MKT scores related to higher ratings, but limited sizes of the effects. Similarly modest effects are observed for the factor score responses.

The test for the MKT effect on the capsule rating also suggests a weak relationship between the MKT scores, having a p-value of 0.10. It also has the smallest intra-district correlation of all the analyses which may suggest less of an effect of the district (or observers) on this measurement in contrast to the other traits measured. The follow-up comparisons only suggest a possible difference between the high and low MKT groups. The four synthesis ratings show interesting variation in the ability of MKT scores to explain responses on those labeled groups. For Synthesis 2 (from the labeled group of items labeled "Implementation" in the instrument), the MKT effect has a p-value of 0.50. For the other synthesis ratings (1=Design, 3= Content, and 4=Classroom), marginal evidence of a significant relationship between MKT and the responses is observed with p-values of 0.05, 0.05, and 0.08 respectively. For the Design, Content, and Classroom ratings, there is some evidence of a difference between low and high MKT scores and some evidence for Content of a difference between low and medium MKT scores.

All of the factor scores show some evidence of a relationship between MKT and the observed scores with similar ICCs for all three models. The strongest evidence of an MKT effect is found in factor 1 (Mathematics Content) and then factor 3 (Student Collaboration). Student centeredness is less well explained by MKT scores. Strong evidence of a difference between high and low MKT groups is found in both mathematics content and student collaboration.

Table 2. Summary of results from mixed models for effect of binned MKT scores.

Response[1]	Differences in Responses between MKT Categories			LRT p-value[2]	District ICC[3]	Pairwise Comparisons, Bonferroni adjusted p-values[4]		
	M vs L	H vs L	H vs M			M vs L	H vs L	H vs M
Capsule	0.31	0.63	0.32	0.10	0.12	0.60	0.08	0.57
Synth1	0.48	0.70	0.22	0.05	0.25	0.17	0.05	1.00
Synth2	0.20	0.35	0.15	0.50	0.30	1.00	0.71	1.00
Synth3	0.09	0.67	0.59	0.05	0.30	1.00	0.08	0.08
Synth4	0.28	0.66	0.38	0.08	0.20	0.76	0.07	0.41
Factor1	0.20	0.47	0.27	0.02	0.33	0.44	0.01	0.19
Factor2	0.12	0.42	0.30	0.08	0.32	1.00	0.09	0.25
Factor3	0.15	0.52	0.37	0.03	0.32	1.00	0.03	0.12

[1]Capsule rating modeled as a seven level ordinal response, Synthesis ratings as five level ordinal responses, and Factor scores as quantitative responses. [2]All based on Likelihood ratio tests (LRTs) for the overall binned MKT effect using Chi-square(2) distributions to approximate p-values. [3]The intra-class correlation (ICC) for teachers in the same district is based on variance of District random effect divided by the sum of that variance and the residual variance. Residual variance for the ordinal models is the variance of the underlying latent trait assumed to be 1. [4]P-values for pairwise comparisons adjusted to control family-wise error rates for the three comparisons in each model.

In general, the MKT seems to be related to the ITC-COP responses but only modestly. Low and medium MKT scores provide no evidence of differences in classroom practice responses and only once were high and medium groups found to have any evidence of a difference. The primary differences were between responses in the high and low MKT groups.

This effect was found across most of the aspects of the ITC-COP instrument except the implementation-based synthesis responses. But where statistically significant, the observed size of the effects was not very large.

Discussion

These results show that in a moderately sized study across multiple states, school districts, and curricula, a teacher mathematics capacity measure is related to multiple mathematics teaching dimensions. However, the relationships are not strong enough and ubiquitous enough across dimensions to justify substitution of an assessment of mathematics knowledge for direct classroom observation in characterizing classroom practice. A larger sample size would provide stronger evidence of the relationship between MKT and aspects of the ITC-COP, but one would not expect the magnitude of the effect to vary substantially. The statistical significance of most of the effects is modest at best. For the significant effects, the magnitude is very modest even for aspects that are most clearly explained by MKT. This suggests that there are important aspects of classroom practice that are not captured by paper and pencil assessments of mathematical knowledge for teaching.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, 2nd ed. New York: Wiley.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 389-407.

Bates, D., Maechler, M., &Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39. http://CRAN.R-project.org/package=lme4.

Christensen, R. (2011). ordinal2---Regression Models for Ordinal Data R package version 2011.05-10 http://www.cran.r-project.org/package=ordinal/.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D.A. Groews (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147-164). New York: Macmillan.

Hanushek, E. (1996). A More Complete Picture of School Resource Policies. *Review of Educational Research*, 66, 397-409.

Hill, H. C., & Ball, D.L. (2004). Learning mathematics for teaching: Results from California's Mathematics Professional Development Institutes. *Journal of Research in Mathematics Education*, 35, 330-351.

Hill, H.C., Blunk, M. L., Charalamobus, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., and Ball, D. L. (2008). *Cognition and Instruction*, 26, 430-511.

Hill, H. C., Umland, K., & Litke, E. *Teacher Quality and Quality Teaching*, under review.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematics knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371-406.

Hill, H. C., Shilling, S. G., & Ball, D. L. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. *The Elementary School Journal*, 105, 11-30.

Horizon Research, Inc. (2000). *Validity and reliability information for the LSC Classroom Observation Protocol.* Retrieved July 1, 2011 from http://www.horizon-research.com/LSC/news/cop_validity_2000.pdf.

Leinhardt, G., & Smith, D.A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology, 77*, 247-271.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Pinheiro, J. & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag,

Putnam, R. T., Heaton, R., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *Elementary School Journal*, 93(2), 213-228.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Rockoff, J. E., Jacob, B. A., Kane, T. J. & Staiger, D. O. (2008). Can you recognize an effective teacher when you recruit one? NBER Working Paper 14485. Cambridge, MA: National Bureau of Economic Research.

Sowder, J.T., Phillip, R. A., Armstrong, B. E., & Shappelle, B. P. (1998). *Middle-grade teachers' mathematical knowledge and its relationship to instruction*. Albany, NY: SUNY Press.

Stein, M. K. & Kaufman, J. H. (2010). Selecting and Supporting the Use of Mathematics Curricula at Scale. *American Educational Research Journal*, 47, 663-693.