Identifying Change in Secondary Mathematics Teachers' Pedagogical Content Knowledge

Robert Powers
U. of Northern Colorado
robert.powers@unco.edu

Shandy Hauk
WestEd
shauk@wested.org

Melissa Goss
U. of Northern Colorado
melissa.goss@unco.edu

Like several other research groups, we have been investigating multiple measures for capturing change in middle and high school teachers' mathematical pedagogical content knowledge (PCK). This article reports on results among 14 teachers (of 16 enrolled) who have completed a distance-delivered master's program in mathematics education. The degree program seeks to develop content proficiency, cultural competence, and pedagogical expertise for teaching mathematics. Analysis included pre- and post-program data from classroom observations and written PCK assessment. Results indicate significant changes in curricular content knowledge on the observation instrument and significant changes in discourse knowledge on both the observation instrument and the written assessment. Path analyses suggest teacher discourse knowledge as measured by the written assessments is significantly related to discourse knowledge as measured by the post-program observation.

Key Words: Pedagogical content knowledge, in-service teachers, professional development

Background

In response to the call for advanced professional education accessible to in-service teachers, the Mathematics Teacher Leadership Center (Math TLC), an NSF-funded Mathematics and Science Partnership project, has developed and is researching a distance-delivered master's program in mathematics education. The primary goals of the program are to develop content proficiency, cultural competence, and pedagogical expertise for the teaching of secondary mathematics (grades 6 to 12). To document the development of mathematics teaching expertise, project research investigates the pedagogical content knowledge of participants before and after the master's degree program. This report is the first to include pre- and post-program data for the first cohort of graduates.

Pedagogical content knowledge (PCK) is a construct described by Schulman (1986) and subsequently refined by others. It encompasses the unique collection of discipline-connected knowledge needed for teaching. As PCK has become widely utilized in research on early grades (K-8) teacher development, a model based on "mathematical knowledge for teaching (MKT)" has emerged (Ball, Hill, & Bass, 2005). Many challenges in measuring PCK have been reported (Hill, Ball, & Schilling, 2008) and most framing of MKT includes some algebra and little in the way of proof-based understandings, such as are found in college mathematics. For the purposes of this research, we use an expanded model of PCK, based on the work of Ball and colleagues, which includes algebra and proof-based advanced mathematics. Working from the foundational three components proposed by Ball et al., the model adds a fourth node of knowledge needed for teaching, discourse knowledge (this aspect brings to the modeling of PCK the mathematical semiotics that was part of Shulman's original description). One way of visualizing the model is as a tetrahedron whose base is the MKT model with apex of discourse knowledge (see Figure 1). Our attention has focused on discourse knowledge and the three "edges" connecting it to the components in the MKT model (Hauk, Jackson, & Noblet, 2010). ***Discourse knowledge*** (**DK**) is knowledge about the culturally embedded nature of inquiry and forms of communication in mathematics (both in and out of educational settings). This collection of ways of knowing includes syntactic knowledge, "knowledge of how to conduct inquiry in the discipline" (Grossman, Wilson, & Shulman, 1989, p. 29). ***Curricular content knowledge*** (**CCK**) is substantive knowledge about topics, procedures, and concepts along with a comprehension of the relationships among them and conventions for reading, writing, and speaking them in school curricula. In its most robust form, this part of PCK contributes to what Ma (1999) called "profound understanding of mathematics" (p. 120). ***Anticipatory knowledge*** (**AK**) is an awareness of, and responsiveness to, the diverse ways in which learners may engage with content, processes, and concepts. Part of anticipatory knowledge growth involves "decentering" – building skill in shifting from an ego-centric to an ego-relative view for seeing and communicating about a mathematical idea or way of thinking from the perspective of another (e.g., eliciting, noticing, and responding to student thinking). ***Implementation knowledge*** (**IK**) is about how to enact in the classroom the decisions informed by knowledge of content and teaching along with discourse understandings. This includes adaptive, in-the-moment, shifting according to curricular and socio-cultural contexts.

This paper describes our efforts to gather evidence of PCK using this four-part framework. We report here on our progress to date in addressing the following research questions:

(1) Does teacher-participant PCK differ pre- to post-program as measured by
  (a) an observation instrument?
  (b) a written assessment?
(2) What is the nature of the relationship between PCK as demonstrated reflectively on the written assessment and in practice on the observation instrument?

While acknowledging the limitations of this non-experimental, small-*n* study, it is valuable in building a foundation for larger scale work in the future. Note that the intention is not to make causal claims. Rather, we are in the early work of testing predictive validity for instruments and exploring potential avenues for capturing PCK and documenting change in it.

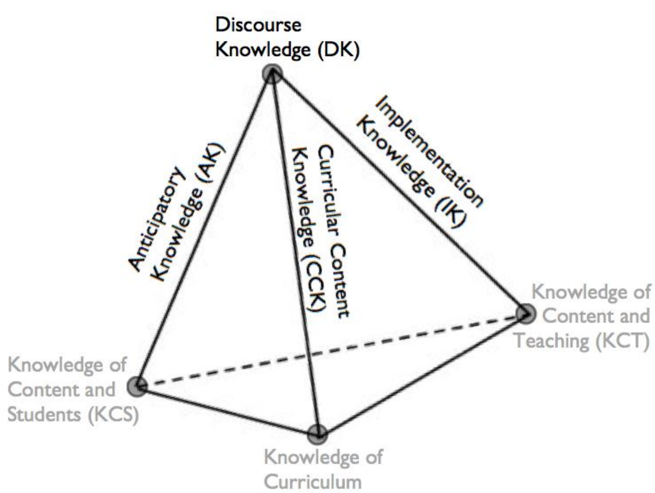

Figure 1. Tetrahedron to visualize relationship among PCK model components. Corners of the base are the aspects of PCK articulated in Hill, Ball, and Schilling (2008).

Methods

Setting: The setting was a blended face-to-face and online delivered master's degree program in mathematics for in-service secondary teachers. Designed to reach urban, suburban, and isolated teachers in rural areas, the program is conducted using a variety of technologies (e.g., *Collaborate* for synchronous class meetings, *Edmodo* for asynchronous communication). Offered through a joint effort at two Rocky Mountain region universities, cohorts of 16 to 20 new students each year complete a 2-year master's program in mathematics with an emphasis in teaching (about half of course credits in mathematics, half in mathematics education).

Participants: Participants for this study were in-service secondary teachers who teach grades 6 to 12 mathematics. Of the 16 who started, 14 completed the coursework of the master's program. All 14 completed the pre- and post-program written assessment. While data included pre-program observations for all 14 teachers, as of this writing there are 10 for whom we have pre- and post-program observations.

Instruments: The development of the written and real-time observation instruments is reported elsewhere (Hauk, Jackson, & Noblet, 2010; Jackson, Rice, & Noblet, 2011). The most important things to note here are that the written assessment included: released items from the LMT (Ball et al., 2008), new items with more complex mathematical ideas modeled on the LMT items, some secondary *Praxis* items, and open-ended extensions to these limited option items. Multi-year test development has included cognitive interviews with in-service teachers and mathematics teacher educators as they did individual items or collections of items. The research team created an alignment of the four PCK constructs of interest across items (e.g., one item might present both curricular content and discourse knowledge challenges while another might foreground curricular content and anticipatory knowledge). These "loadings" of multiple PCK constructs to items is a purposeful part of the non-linear model underpinning test design. Each item on the written test loaded on at least two of the four PCK constructs. Consequently, factor analysis was not appropriate given this confounding of variables. In addition to the established face validity of the tests, tests of the constructs' internal consistency (Cronbach's alpha) indicate, for the pre-test, good overall reliability ($\alpha = .81$), good reliability on CCK ($\alpha = .81$), acceptable reliability on DK ($\alpha =.76$), and marginal reliability on AK ($\alpha = .55$). The PCK post-test had acceptable reliability overall ($\alpha = .75$), acceptable reliability on CCK ($\alpha = .75$), and DK ($\alpha = .73$) (George & Mallery, 2003). However due to an unacceptably low reliability on the post-test for AK for this first cohort of teachers ($\alpha < .5$), we cannot deem the written PCK tests as validly measuring anticipatory knowledge for the group. The observation instrument, based on the LMT video observation protocol (see LMT website; development reported elsewhere) showed good reliability overall ($\alpha = .85$); good reliability on CCK ($\alpha = .84$), DK ($\alpha = .89$), and IK ($\alpha = .85$); and acceptable reliability on AK ($\alpha = .78$). Like the LMT video protocol, the observation tool used samples (6 minutes each: 3 minutes observed, 3 minutes to identify presence/absence of each protocol category in the observed segment; each class visit had 7 to 12 segments). An "observation" was three consecutive classroom visits.

We did pre-program classroom observations in the spring term prior to teachers entering their first course of the master's program. Post-program observations were in the spring term two semesters after the teacher completed the program. Each included three separate classroom visits by the same researcher(s). Experienced observers trained new observers to use the instrument; new raters practiced using the protocol on video data, conducted their first observations of teachers in tandem with an experienced observer, and team members met to calibrate ratings and reconcile disagreements. Inter-rater reliabilities were greater than 0.8.

Teacher-participants completed the written pre-test at the beginning of their first class session in the program. Of the 14 teacher-participants who completed the program, 9 completed the post-program written test at the program closure meeting. For the 5 unable to attend the meeting, members of the research team administered the test at the teachers' school of employment. For each administration of the test, members of the research team created answer keys for multiple-choice items and a scoring rubric for short answer items. The rubrics were informed both by expected responses identified by item developers as well as cognitive interview data. The procedure for developing the rubric was (1) write a desired response, (2) list other anticipated responses, (3) read the responses from a subsample of participants, (4) come to consensus on a scoring rubric. Two or more research team members scored tests separately, compared scores, and met to reconcile any disagreements.

To date the research team has observed 10 teachers after completion of the program. The counts for each of the observation variables were summed and divided by the number of segments observed to report a relative frequency for each variable for each teacher. A teacher having a score of 23.25 on "Explicit Talk about Math" means that the rater(s) noted the teacher exhibiting explicit talk about mathematics during 23.25% of the segments observed. Similarly, on the written test, researchers calculated relative frequency percent scores for each of the four PCK constructs by summing teacher scores on items coded for the construct and taking the percent out of total points possible on each construct. To answer the research question of the impact of the master's program on teachers' PCK, we compared entrance and exit data from the written assessment and the observations using paired-samples *t*-tests.

To model the relationship between teachers' PCK as measured with the written items and in practice as observed, we conducted a path analysis on each of the four PCK constructs. The model considered the pre-test and pre-observation scores as exogenous variables. Thinking that change in knowledge leads to change in action, the model examined the effects of the exogenous variables (pre-scores) on the written post-test; then examined the effects of those three variables on the post-program observation scores. The results report path analysis for CCK and DK (the AK construct was not robustly reliable and the written test did not measure IK).

Results

Observations

Table 1 (see appendix) gives information on pre- and post-program observations for the 10 teachers for whom complete data are available. The table presents the means, standard deviations, differences from pre- to post-program, and results of paired samples *t*-tests on each variable. Because of the number of statistical analyses performed, a cutoff *p* value of 0.0015 (rather than 0.05) is appropriate, based on a Bonferroni correction (Bland & Altman, 1995).

With this threshold for alpha, there are two statistically significant results. One was in the observation category "General language for expressing mathematical ideas (overall care and precision with language)." While such use of general language was seen, on average, in about 49% of pre-program classroom segments, by the end of the program it was present in more than 80% (M=80.34, SD=19.71). The other significant result was in the category "Mathematical descriptions (of steps)" (i.e., segments where the teacher or students accurately used explicit language to describe the steps of some process). On average, across pre-program observations, this was seen in about 40% of class segments (M=40.28, SD=21.94), increasing to almost 70% of the time, post program (M=68.10, SD=19.31). Three other observed variables appear to be approaching significance (i.e., p<.01): the percent of segments where (a) student voices were present in the room (increasing from 80% to 90% of segments), (b) teachers were observed to use conventional notation (increasing from 54% to 90% of segments), and (c) fewer mathematical errors occurred (decreasing from about 4% of the time to nearly 0%).

Table 2 presents the results of aggregating observation variables associated with each of the PCK constructs. Based on the Bonferroni correction, none of the results were statistically significant. Two approached significance: curricular content knowledge (increasing from 45% to 57% of segments) and discourse knowledge use (increasing from 48% to 61% of segments).

Table 2. Paired samples t*-tests for PCK Constructs from Observation Instrument*

	Pre-program (*N*=10)		Post-program (*N*=10)			
PCK Construct	*M*	*SD*	*M*	*SD*	*t*	*p*
Curricular Content Knowledge (CCK)	45.12	13.18	56.64	10.66	4.31	.002@
Discourse Knowledge (DK)	48.25	13.47	61.27	10.27	3.92	.004@
Anticipatory Knowledge (AK)	44.18	12.97	54.26	17.56	1.95	.083
Implementation Knowledge (IK)	59.16	15.13	66.12	9.97	1.54	.159

@ indicates approaching significance, with a $p < .015$

PCK Observations and PCK Test

We conducted a variety of path analyses on the data. In the presentation we focus on the analysis of the Discourse Knowledge (DK) construct, examining the relationships and potential predictive power of the written instrument. Figure 1 shows the full model for discourse knowledge. There was a significant effect of the pre-test (β=.78, *SE*=.26, p<.05) and no significant effect of the pre-observation on the DK post-test. There was no significant effect of either the pre-observation or pre-test on the post-observation DK score, although, like CCK, the effect of the pre-test was negative (β=-.58, *SE*=.19). Finally, there was a significant effect of the post-test on the DK post-observation (β=.92, *SE*=.17, $p < .05$).

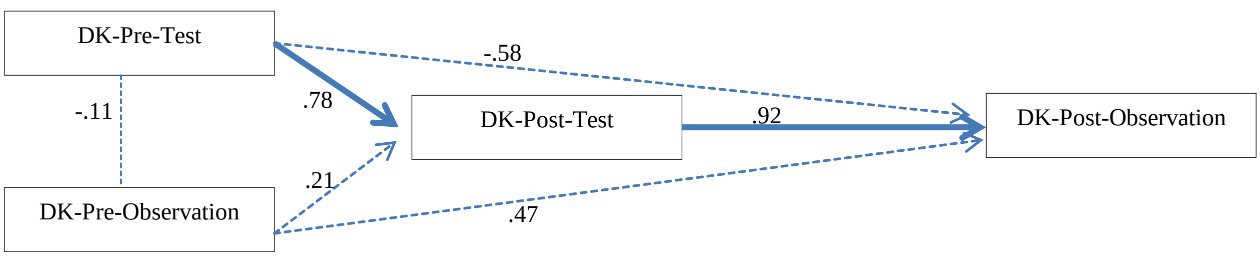

Figure 2. Path diagram for discourse knowledge (DK) construct.

Discussion/Applications/Implications

Because of the small sample size, the study is underpowered for full validation of the assessment and observation scores. Additionally, the small sample size makes generalizing the results problematic. What is apparent is that pre- to post-program written test score changes suggested positive potential outcomes of the master's program in the target area of investigation: development of pedagogical expertise for teaching secondary mathematics, particularly in the communication skills of responsive classroom discourse. The significant increase in curricular content knowledge (CCK) from pre- to post-program teacher observations may reflect the master's program emphasis on increasing participant understanding of advanced mathematics and deepening secondary-school-level-appropriate conceptual connections. This is evident in some of the significant increases on individual variables in categories. For example, the significant increase in the use of conventional notation may indicate that the master's program supported teacher-participants in the habit of using conventional notation to communicate. In addition, the mathematics courses required participants to be explicit about their thinking, reasoning, and justification of answers, which may help explain the significant increase in mathematical descriptions category. However, there were no significant increases in the mathematical explanations or the mathematical justification of the reasoning process, so more work needs to be done in the program to support teachers' attention in these important realms of mathematics teaching and learning (perhaps as they challenge the prescribed curricula, which tend *not* to foreground these things). Finally, the reduction in observed errors may indicate a stronger content knowledge for teaching secondary mathematics.

The significant increase in discourse knowledge (DK) on the written test and in observations may indicate the effectiveness of the master's program mathematics education courses. In particular, the program's emphasis on mathematics pedagogy that made explicit the research-based evidence of student-centered classrooms that support the construction of knowledge of students rather than the transmission of knowledge by teachers. For example, observers saw significant increases in the percent of small group work and in students' voices in the classroom. This may indicate that the teachers' practice shifted to a decentered (or some forms of "learner-centered") approach. Additionally, the program included several credit hours of reading and writing about mathematics education research focused on the NCTM process standards. There was a concomitant significant increase in teachers' explicit talk about reasoning. Finally, the increase in discourse knowledge in general may be attributed to the pedagogy courses that allowed participants to read research and experience what good mathematics discourse "sounds like and feels like" (Cohort 1 participant, personal interview, October 8, 2012).

The path analyses relating PCK as demonstrated on the written test and in practice provide interesting results that need further investigation. As noted, the path diagram for discourse knowledge, Figure 1, suggests that the written test may have predictive value in capturing classroom practice. If this turns out to be a robust result, across populations of teachers, it could reduce or eliminate the need for expensive classroom visits when attempting to determine impact on practice. Researchers need to conduct further investigation into the ways to measure these constructs and to extend the research to larger, more generalizable samples to verify these results. Additionally, researchers need to investigate the negative, albeit not significant, direct effect of the pre-test on the post-observation.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. DUE0832026. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

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

Bland, J. M., Altman D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ* [Note: Formerly *British Medical Journal*] 310(6973),170.

George, D. & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.*11.0 update (4th ed.). Boston: Allyn & Bacon.

Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 23-36). UK: Pergamon.

Hauk, S., Jackson, B., & Noblet, K. (2010). No teacher left behind: Assessment of secondary mathematics teachers' pedagogical content knowledge. *Proceedings for the 13th*

Conference on Research in Undergraduate Mathematics Education (electronic). PDF available at sigmaa.maa.org/rume/crume2010/Archive/HaukNTLB2010_LONG.pdf

Hill, H., Ball, D., & Shilling, S. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372-400.

Jackson, B., Rice, L., & Noblet, K. (2011). What do we see? Real time assessment of middle and secondary teachers' pedagogical content knowledge. In S. Brown (Ed.) *Proceedings for the 14th Conference on Research in Undergraduate Mathematics Education (electronic.)* PDF available at sigma.maa.org/rume/crume2011/Archive/

Learning Mathematics for Teaching (LMT) Website: http://sitemaker.umich.edu/lmt/home

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahway, NJ: Erlbaum.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Table 1. Paired Samples t*-tests for Observation Variables.*

	Pre-program (*N*=10)		Post-program (*N*=10)			
Observation Item	*M*	*SD*	*M*	*SD*	*t*	*p*
Format for Segment						
Whole Group	51.61	22.16	63.39	18.40	1.170	.272
Small Group	23.79	22.51	39.26	22.51	2.832	.020
Individual	41.70	26.03	28.98	12.15	-1.610	.142
Lesson/Segment Type						
Review	26.46	18.03	22.46	14.17	-0.549	.596
Introducing tasks	7.23	4.74	10.64	5.31	2.262	.050
Student work time	45.00	24.16	50.04	16.76	0.586	.572
Direct instruction	24.15	15.27	33.00	16.53	1.201	.260
Synthesis or closure	5.77	4.91	8.10	6.02	1.147	.281
Math Teaching Practices						
Voices – Students	79.82	18.32	89.29	15.88	3.375	.008@
Voices – Teacher	80.77	21.98	93.81	8.23	1.949	.083
Real-world Problems	26.55	28.47	36.50	32.44	.826	.430
Interprets Students' Work	63.33	18.39	73.01	13.75	2.120	.063
Explicit about Tasks	82.20	16.20	87.52	12.42	.916	.384
Explicit Talk about Math	59.03	27.18	75.59	11.91	1.801	.105
Explicit Talk about Reasoning	29.93	23.18	49.48	17.94	2.821	.020
Instruction Time	86.10	10.20	87.02	6.81	.249	.809
Encourages Competencies	67.07	26.83	45.04	40.52	-1.420	.189
Knowledge of Math Terrain						
Conventional Notation	54.39	21.38	79.95	15.25	3.353	.008@
Technical Language	72.59	17.38	77.67	13.53	.760	.467
General Language	49.06	13.88	80.34	19.71	4.528	.001*
Selection for Ideas	87.17	8.31	91.16	5.68	1.989	.078
Selection to Represent Ideas	31.70	23.97	43.64	25.51	1.892	.091
Multiple Models	17.80	14.90	33.69	24.20	2.138	.061
Records Work	59.67	28.25	52.01	20.20	-.585	.573
Math Descriptions	40.28	21.94	68.10	19.31	5.003	.001*
Math Explanations	40.65	23.26	55.80	16.29	1.782	.108
Math Justification	14.32	16.13	23.09	11.05	1.928	.086
Math Development	84.50	16.57	88.67	6.11	.753	.471
Errors – Not Present	96.27	2.67	99.78	.69	3.858	.004@

@ $p < .015$, * $p < .0015$