WHAT DO WE SEE? REAL TIME ASSESSMENT OF MIDDLE AND SECONDARY TEACHERS' PEDAGOGICAL CONTENT KNOWLEDGE

Billy Jackson	Lisa Rice	Kristin Noblet
Saint Xavier University	University of Wyoming	U. of Northern Colorado
bjackson@sxu.edu	lrice4@uwyo.edu	kristin.noblet@unco.edu

The article reviews efforts to develop an observation protocol to assess the pedagogical content knowledge (PCK) that middle and high school teachers may develop and demonstrate in the classroom over time as part of their participation in a master's program for secondary mathematics teachers. We observed each of 16 teachers in real time using the instrument, before involvement in the project and again during the second semester of participation. Aspects of the protocol measure four critical components of PCK including curricular content, discourse, anticipatory, and implementation knowledge. We present a quantitative analysis of the observations and discuss various challenges faced in the instrument development and its relation to similar protocols used by others previously.

Key Words: pedagogical content knowledge, discourse, inter-rater reliability, teaching moves

There have been several approaches to measuring the pedagogical content knowledge (PCK) of practicing teachers. Indeed, Hill, Ball, and Schilling (2008) and Hauk, Jackson, and Noblet (2010) have documented their development of written instruments designed to assess aspects of PCK. Both groups have developed theoretical frameworks for PCK that have similarities and some differences. One of the principle differences is that the Hill, Ball and Schilling model seeks to measure each of their proposed categories of PCK as distinct from each other, while Hauk, Jackson, and Noblet take a non-linear approach that presumes measurement overlap among categories. To understand what a non-linear model means, PCK can be thought of as a dynamical system as its development in teachers progresses over time. Hill, Ball, and Schilling presume that for their written instrument, a teacher's PCK can be measured by taking a linear combination of items that each pertain to exactly one of their three PCK constructs: knowledge of content and students (KCS), knowledge of content and teaching (KCT), and knowledge of content and curriculum (KCC). In contrast, a single item on the written instrument from Hauk, Jackson, and Noblet can touch on multiple facets of their typology of PCK, and hence introduces overlap in these facets, creating nonlinearity in the system. While Ball and colleagues acknowledge that the relationship between a teacher's KCS, KCC, and KCT may well be non-linear (LMT, 2006), the current model makes the linearity assumption. A mathematical model of PCK with consideration of Hauk, Jackson, and Noblet's approach will require a non-linear system.

Hauk, Jackson, and Noblet (2010) discuss PCK in terms of four components: curricular content, discourse, anticipatory, and implementation (action) knowledge. *Curricular content knowledge* is "substantive knowledge about topics, procedures, and concepts along with a comprehension of the relationships among them *as they are offered in school curricula*" (p. 2). *Discourse knowledge* is substantive knowledge "about the culturally embedded nature of inquiry and forms of communication in mathematics (both in and out of educational settings" (p.2), and as such includes knowledge of mathematical syntax as a sub-category. *Anticipatory knowledge* "is an awareness of, and responsiveness to, the diverse ways in which learners may engage with content, processes, and concepts" (p. 3). *Implementation* or *action* knowledge "includes knowledge about how to adapt teaching according to content and socio-cultural context and *enact in the classroom* the decisions informed by content, discourse, and anticipatory understandings" (p. 3).

Both groups' written assessments use multiple-choice items and are limited in measurement of teacher knowledge related to implementation. Such knowledge is more challenging to assess as this type of knowledge requires actions executed in the classroom (e.g., teacher moves). That is, the written assessments could not test for this type of knowledge because it requires that the teacher act "in the moment." At best, any written item could only gauge what a teacher *might* do in certain situations (e.g., see Ball, Hill, & Schilling, 2008). In order to validate their written instrument, Ball and others (Learning Mathematics for Teaching (LMT), 2006) developed another instrument aimed at quantitatively measuring aspects of elementary and middle school teachers' classroom practice. Ten K-8 teachers who had taken a PCK test were videotaped three times prior to, during, and after participation in professional development. Over the course of a year, a team of mathematicians, mathematics educators, mathematics teachers, and non-specialists analyzed the videos for various aspects of mathematics and mathematics teaching present in each lesson. A rubric was developed containing several items and video reviewers trained for and then coded each 5 minute segment of each lesson for different categories of teacher move or classroom interaction. Each category had four possible codes: Present and Appropriate (PA), Present and Inappropriate (PI), Not Present and Appropriate (NPA) and Not Present and Inappropriate (NPI). LMT team leaders noticed early on a wide variability in how individuals coded lessons based upon the individuals' own professional backgrounds, and so to help ensure inter-rater reliability, the lessons were all recoded in pairs. A glossary describing each category (column) in the observation rubric was written, with each description giving some detail on when each code should be assigned during a segment.

Theoretical Perspective

Our research blends the Hauk et al., framework for PCK and the LMT observation instrument designed by the research team at the University of Michigan. We take the view that the teacher actions or moves (or the absence thereof) in the LMT protocol can be observed in the classroom, and that such actions or moves can be described (at least approximately) in a predetermined coding format *independently of the researcher involved*. Now, this is not to say that two different researchers may not observe and record different things (as frequently happened with the team at the University of Michigan and for our team) for a given segment a few minutes long, but, like the LMT tool, for an entire class period observation we would expect variation between observer totals to be minimal.

We use here the typologies of Hauk, Jackson, and Noblet (2010). The reason is that any particular move that a teacher makes in the classroom can demonstrate multiple facets of PCK simultaneously, and hence we take their view that the strands of PCK are interwoven during instruction. Also, Hauk, Jackson, and Noblet make cultures in the classroom an explicit part of their definitions, which in turn may be part of teacher decisions to make certain moves in response to them.

Discourse knowledge requires some further investigation here, as it is necessary to make explicit how we are defining it. Ryve (2011) points out that it is indeed important for authors to provide definitions and epistemologies when discussing discourse as "some key words or phrases are particularly complex and multifaceted" (p. 167). To this end, we use Gee's (1990) definition of Discourse (with a capital 'D'), which he distinguishes from discourse (with a lower case 'd').

For Gee, “Discourses are ways of being in the world, or forms of life which integrate words, acts, values, beliefs, attitudes, social identities, as well as gestures, glances, body positions and clothes” (pp. 142-143). However, “discourse” with a little “d” refers to “connected stretches of language that makes sense” (pp. 142-143). Hence for Gee, discourse is a part of Discourse. Note that if discourse is "connected stretches of language that make sense" to those taking in and/or producing the connected stretches, then semantics and syntax are part of discourse. That is, discourse could be thought of to have at least three components: what is intended, what is perceived, and the associated interaction that allows the exchange from intender to perceiver. Such discourse applies to written/read and or spoken/heard interaction. For the purposes of this paper, whenever the word “discourse” is used, we will mean the word in the broader sense of Gee’s “Discourse” as such things as gestures will be considered discourse for our purposes, though, admittedly, here usually we will refer to spoken language. The discourse used here is that of mathematical discourse rather than that of what Ryve (2011) terms “general educational discourses” (p. 174) that can be relevant to mathematical discourse but are nonmathematical in nature.

Ryve also points out that our epistemological assumptions have consequences for how we conceptualize mathematics and learning. In light of this, we take Sfard’s (2006, 2008) view that mathematical objects are concrete and tangible and hence can be perceived by the learner. Examination of the LMT (2006) instrument and the corresponding glossary shows that Ball and others share this same view in their work as well.

Discourse knowledge is perhaps the most unique type of knowledge among the classifications of Hauk, Jackson, and Noblet. A comparison of discourse knowledge to curricular content knowledge reveals that curricular content can become quite static at some point while discourse knowledge constantly evolves due to the nature of discourse being imbedded in the culture of the classroom, which is subject to change on both short and long time scales. In fact, many shifts in PCK that occur as a teacher develops it can be attributed to changes in discourse knowledge, since discourse is woven into the other three components very deeply (i.e. there is substantial overlap between discourse and the other three). Discourse knowledge, like implementation knowledge, might have aspects of social co-construction whereas curricular content and anticipatory knowledge may tend to be individually constructed (though such construction can be socially mediated).

The research questions for the work reported here are: How might we track the effects of professional development through changes in observed PCK? If traceable, how might professional development be designed to foster particular classroom moves through changes in PCK? Work on both of these questions continues, and we will primarily address the first here but some attention will be given to implications of current results for the second.

Methods

We built our work from a careful consideration of the LMT work. The LMT group point out in their technical report that there is a need to develop an instrument for doing observations in real time (LMT, 2006, p. 20). In order to address this need, we examined their observation protocol in some depth and determined which items were most appropriate given our focus on observing in secondary mathematics classrooms in real time (the LMT work was in grades K-8). A copy of our protocol is in the appendix. Their protocol contained over 30 categories. To streamline for real-time observation we shortened to a protocol containing 20 items. Some of their categories were replaced or condensed in our version. For example, in the LMT version, the researchers created columns for the following: selection of correct manipulatives, and other visual and concrete models to represent mathematical ideas (their column II.e on sheet 2) and multiple models (column II.f on sheet 2). In our version, these two columns were condensed into the column that we titled multiple representations, which could include all of the things that the LMT team was looking for in II.e and II.f.

Great care was taken in finding an appropriate length of a segment to be viewed during the class. The team started with the 5 minute length that the LMT used for recorded sessions, but it soon became clear that a "5 minute on, 5 minute off" strategy in which the researcher would observe for 5 minutes and then record tallies on the protocol during the next 5 minute interval would result in 5 or fewer codings per class period for each category (not to mention the fact that it was challenging to hold five minutes of observation in one's head while attempting to record everything about it). Eventually, the team agreed upon observing for 3 minutes, and then recording for 3 minutes. Thus, it took 6 minutes to capture a particular 3 minute chunk of observation on the observation sheet, which meant about 9 samples in a short class period (45 to 55 minutes) and up to 15 in an extended or block period.

After the team started using the protocol, we began to reexamine the glossary that the LMT team had developed. We found that trying to use the instrument in real time created new challenges with respect to inter-rater reliability. In particular, the words "explicit" and "inappropriate" leave much room for interpretation even in the definitions provided by the LMT team. Though we used many of the same column categories and indentifying language as they did, we also saw it was important to craft definitions and create a new glossary. The idea was to create an instrument with sufficient examples and non-examples for each category that it could act as a coding book: a guide to the intended viewpoint of the protocol and how to observe through a particular lens. The eventual goal is to have an instrument that is terse but of sufficient detail that individuals can observe classrooms after a brief calibration training paired with a practiced observer or after watching video of the same.

For example, while our glossary continues to be refined, we felt a need to be, well, more explicit about what "explicit talk about a topic or subject" meant. Currently, our glossary description of this category is: any utterance from student or teacher in which a topic or subject is stated verbally or in writing or by reference to a clear written or verbal precursor familiar to people in the room (e.g. through gestures). In vivo exemplars have been included in our glossary to demonstrate categories. For example, during one 3-minute segment, the teacher presented the Fundamental Theorem of Algebra. The exercise the teacher assigned called for students to find a polynomial of lowest degree with real coefficients that had certain prescribed roots. At one point, an exercise asked for a polynomial with roots $3i$, 4, and 5. The teacher produced a monic degree 4 polynomial with these 3 prescribed roots, and a student asked why it was necessary to have $-3i$ as a root when this number was not contained in the list. The teacher responded that since $3i$ was a root, its conjugate $-3i$ also had to be a root. The student again asked why this must be true when $-3i$ was not listed, and the teacher replied "because conjugates are always roots." The researcher coded this particular segment as NPI in the "explicit talk about ways of reasoning" column due to the teacher's not addressing directly the student's question in a way that appeared to make sense to the student. The segment was also coded as NPI in the "interprets students' productions/student errors" column due to the fact that in the glossary, this column includes the teacher listening to what the student says and responding to the student.

Each column of the protocol was assigned a quadruplet of the form (c, d, a, i) where the values of c, d, a, i were determined by the research team based on the descriptions of the categories **C**urricular Content, **D**iscourse, **A**nticipatory, **I**mplementation and the glossary description of the category represented by the column. The value for *c, d, a,* or *i* was 1 if a particular kind of knowledge was present in the observable category, or 0 otherwise. That is, we would say a particular category "loaded" on a particular PCK aspect if the appropriate presence of that category suggested that the teacher demonstrated that component of PCK during the segment observed. Research team members spent a significant amount of time on coming to a consensus on each of the loadings, particularly implementation knowledge. Working to understand how in what ways the observation categories were aligned to the idea of action knowledge being demonstrated in the classroom. One challenge in defining this particular type of knowledge is that the other three are interwoven with it so much that at times it can be difficult to "tease apart" implementation knowledge from say anticipatory knowledge. After much discussion, we began to understand that implementation knowledge had to meet **both** criteria given in the definition by Hauk, Jackson, and Noblet (i.e., adapting and enacting). This categorical inductive coding left some of the columns without non-zero alignment to any PCK codes (e.g. instructional time spent on mathematics(>75%)). One particularly interesting column titled "encourages diverse mathematical competencies" has a unique feature: we determined that this column loaded heavily on PCK by assigning it a quadruple of (1,0,1,1). The category draws on curriculum content knowledge because the teacher will demonstrate some knowledge of the curriculum even if peripherally since the teacher must be informed about the curriculum enough to know that she is encouraging diverse mathematical competencies. She is also demonstrating the use of her anticipatory knowledge when the column is checked as PA because she must be aware of how the students may interact with the material at hand in order to encourage the competencies. She adapts her teaching according to the diversities that arise and makes choices in her instruction accordingly, and so the observer will see her use her discourse knowledge during the segment.

The columns that have discourse knowledge present share a feature the definitions of both take into consideration the *interaction* that occurs between the teacher and students as opposed to solely the *actions* of the teacher. Hence, another difference between the typologies of Ball, Hill, and Schilling and Hauk, Jackson, and Noblet becomes evident. The approach of the latter attends to the relational and interactional. Thus, our instrument and the lens it takes may allow for the researcher to look for certain aspects of the classroom that the LMT instrument does not.

Results

The research team observed two cohorts of teacher participants (TP) for a masters program for teachers (n=19 and n=16 respectively). We observed both cohorts before entering the program and the first cohort during their second semester in the program. A frequency analysis was performed for each observation and t-tests performed.

The frequency analysis began by summing the PA's and PI/NPI's for each column. We then subtracted the number of inappropriates from the number of appropriates, assigning a value of +1 for each segment checked PA, -1 for each segment checked PI/NPI, and a value of 0 for each segment marked NPA. After performing this tally for each column, each of the column totals in the protocol were normed in each observation to account for varying numbers of segments coded among the 2 to 3 observations done for each teacher. A PCK score for each of curricular content knowledge (CCK), discourse knowledge (DK), anticipatory knowledge (AK) and implementation knowledge (IK) was assigned by summing the columns for which the team assigned a value of 1 in the quadruplet. For example, the column "encourages diverse mathematical competencies'' contributes to the CCK, AK, and IK scores but not to the DK score since the column was assigned a quadruple of (1,0,1,1) as explained in the previous section. The normed PCK scores from each observation were then averaged and the teacher was assigned that average in each component for pre and follow up observations.

A paired *t*-test was then done for the first cohort on the pre and follow up observations, and an independent samples t-test was done for comparison between the first and seconds cohorts' pre-observations. The results are summarized in the tables below:

Type of Knowledge	*Mean*	*p-value*
CCK, pre, cohort 1	.5926	
CCK, follow up, cohort 1	.5867	.942
DK, pre, cohort 1	.4360	
DK, follow up, cohort 1	.4787	.394
AK, pre, cohort 1	.6239	
AK, follow up, cohort 1	.5098	.03
IK, pre, cohort 1	.5188	
IK, follow up, cohort 1	.5051	.774

Type of Knowledge	Mean	p-value
CCK, pre, cohort 1	.593	
CCK, pre, cohort 2	.605	.818
DK, pre, cohort 1	.436	
DK, pre, cohort 2	.6214	.000
AK, pre, cohort 1	.624	
AK, pre, cohort 2	.588	.459
IK, pre, cohort 1	.519	
IK, pre, cohort 2	.588	.081

Cohort 1 showed no significant differences in the pre and follow up observations in any of the categories except for anticipatory knowledge, in which a significant decrease was observed. This decrease could be due to any number of factors, including an implementation dip. Comparing cohorts 1 and 2, we see that cohort 2 had statistically significantly different DK and IK scores on arriving in the master's program. To understand what these means represent, we first note that inappropriates were assigned infrequently during the observations as researchers tended to describe the classes as average to above average. Thus, a mean of .519 for cohort 1 IK implies that the teachers in cohort 1 demonstrated the use of their IK in about 51.9% of the segments observed, while cohort 2 demonstrated the use of their IK in about 58.8% of the segments observed. It is important to note that this does not mean that cohort 2 has more IK than cohort 1 as a teacher can have knowledge but not use this knowledge during any particular instructional segment. The statistical differences along with other external factors could suggest that the two cohorts are different as populations.

It is interesting to note the difference in discourse knowledge between the to cohorts. It is possible that the discourse knowledge is a factor in the difference in the implementation knowledge between the two cohorts due to the overlap between these two components of PCK. One possible implication of this is that designers of professional development (PD) who wish to create shifts in PCK may be effective by targeting teachers' discourse knowledge. Hence, for example, to advance a teacher's anticipatory knowledge, a PD designer may create sessions designed to build upon her discourse knowledge since these categories overlap in PCK framework.

Conclusions

The model for PCK offered by Hauk, Jackson, and Noblet has some striking differences from that of Hill, Ball, and Schilling. The instrument our research team developed based upon the LMT project in many ways highlights many of those differences. Perhaps one of the biggest strengths of using the nonlinear model is that it foregrounds the importance of discourse in the mathematics classroom, activity that creates interaction among the teacher and students rather than being a sole action of the teacher. It is also possible that discourse knowledge may be responsible for shifts in PCK due to its intrinsic overlap with the other components that Hauk, Jackson, and Noblet propose. This realization could have consequences for the design and implementation of professional development aimed at targeting PCK. We plan further research in this area to examine how PD can be designed with explicit attention to discourse and ways to "unpack" what it means to talk about Discourse (and discourse) in the classroom, particularly in classrooms where students and teachers from myriad cultural Discourse communities interact.

Acknowledgements

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

Gee, J. P. (1990). *Social linguistics and literacies: Ideology in discourses* (1st ed). London: Falmer Press.

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

Learning Mathematics for Teaching (2006). *A Coding rubric for Measuring the Mathematical Quality of Instruction* (Technical Report LMT1.06*)*. Ann Arbor, MI: University of Michigan, School of Education.

Ryve, A. (2011). Discourse research in mathematics education: A critical evaluation of 108 Journal Articles. *Journal for Research in Mathematics Education,* 42, 167-198.

Sfard, A. (2006). Participationist discourse on mathematics learning. In J. Maasz & W. Schloeglman (Eds.), *New mathematics education research and practice* (pp. 153-170). Rotterdam, The Netherlands: Sense.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing.* Cambridge, UK: Cambridge University Press.

Appendix-Observation Protocol

Math TLC Teaching Observations: Recording Sheet

Teacher | **Course** | **Date** | **Time** | **OBS #**

Directions: In all but A., B., and C.a., choose only one option for each class segment.

I. Instructional Format and Practices

	A. Format for segment			B. Lesson/segment type					C. Mathematics Teaching Practices																															
	a. whole group	b. small group/ partner	c. individual	a. review, warm up or homework	b. Introducing major task	c. Student work time	d. Direct instruction	e. Synthesis or closure	a. Voices in the classroom (students, teachers, or both)				b. Real-world problems or examples				c. Interprets students' productions/ students' errors				d. Explicit about student tasks, work, and success				e. Explicit talk about mathematical practices & language				f. Explicit talk about ways of reasoning				g. Instructional time is spent on mathematics (>75% of segment)				h. Encourages diverse mathematical competencies			
									Students		Teachers		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP	
Segment									A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I
1																																								
2																																								
3																																								
4																																								
5																																								
6																																								
7																																								
8																																								
9																																								
10																																								
11																																								
12																																								
13																																								
14																																								
15																																								
16																																								
17																																								
18																																								
19																																								
20																																								

Math TLC- Teaching Observation Recording Sheet

II. Knowledge of mathematical terrain of enacted lesson

DATE: ________

	a. Conventional notation (mathematical symbols)				b. Technical language (mathematical terms and concepts)				c. General language for expressing mathematical ideas (overall care and precision with language)				d. Selection of numbers, cases & contexts for mathematical ideas				e. Uses multiple ways to represent mathematical ideas				f. Makes links/connections among mutliple representations of mathematical ideas				g. Publicly records the mathematics for the class				h. Mathematical descriptions (of steps)				i. Mathematical explanations (giving mathematical meaning to ideas or procedures)				j. Mathematical justifications				k. Development of mathematical elements of the work (i.e., moving the math along)				l. Computational errors or other mathematical oversights			
	P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP		P		NP	
Segment	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I	A	I
1																																																
2																																																
3																																																
4																																																
5																																																
6																																																
7																																																
8																																																
9																																																
10																																																
11																																																
12																																																
13																																																
14																																																
15																																																
16																																																
17																																																
18																																																
19																																																
20																																																

Overall level of teacher's knowledge of mathematics

Low	Medium	High

Math TLC- Teaching Observation Recording Sheet