The Veritable Quandary of Measuring Teacher Content Knowledge in a Math and Science Partnership

Carole Basile, Ed.D.
Co/PI, Co-Director, Rocky Mountain Middle School Math and Science Partnership
Director, Center of Applied Math and Science for Innovation and Competitiveness
University of Colorado at Denver and Health Sciences Center

Karen Koellner, Ph.D.
Associate Professor, Mathematics Education
School of Education and Human Development
University of Colorado at Denver and Health Sciences Center

Doris Kimbrough, Ph.D
PI/Co-Director, Rocky Mountain Middle School Math and Science Partnership
University of Colorado at Denver and Health Sciences Center

Michael Jacobson, Ph.D.
Co-PI, Rocky Mountain Middle School Math and Science Partnership
Chair, Mathematics Department
College of Liberal Arts and Sciences
University of Colorado at Denver and Health Sciences Center

Linda Morris
Co-PI, Rocky Mountain Middle School Math and Science Partnership
Science Director
Jefferson County Public Schools

Barbara Heath, Ph.D.
Evaluator, Rocky Mountain Middle School Math and Science Partnership
East Main Educational Consulting, LLC

Aruna Lakshmanan, Ph.D.
Statistician, Rocky Mountain Middle School Math and Science Partnership

The Veritable Quandary of Measuring Teacher Content Knowledge

There is renewed attention on the content knowledge that middle school teachers need to know, especially in mathematics and science. The Principles and Standards for School Mathematics document suggests that "teachers must know and understand deeply the mathematics they are teaching and be able to draw on that knowledge with flexibility in their teaching tasks" (NCTM, 2000, p. 17). There is substantial evidence that teachers typically do not hold this rich and connected knowledge of mathematics or science, nor do they teach in ways that are consistent with the national standards (Jacobs, 2006; Mewborn, 2003). In order to carry out the demands of reform-oriented mathematics instruction, teachers need increased opportunities to broaden and deepen their professional knowledge. In particular, they need opportunities to gain knowledge of mathematics for teaching, including both specialized subject matter knowledge and pedagogical content knowledge (Ball, Hill, & Bass, 2005). In a similar vein, the National Science Standards state that "Science has a rapidly changing knowledge base and expanding relevance to societal issues, and teachers will need ongoing opportunities to build their understanding and ability" (p. 55). In addition, Loucks-Horsley, Hewson, Love, Stiles (1998) call for professional development that concentrates on specific issues of science content that are derived from research and build on teachers' prior knowledge.

The call for an extended knowledge base has brought awareness to licensure and certification requirements that vary among states. Certification for states runs along a continuum that includes K-8 certification, middle level certification, and secondary certification that is used for teachers who teach 7th through 12th grade. To this end, debates have ensued about the appropriate requirements for middle school mathematics and science teachers given the different types of certifications. This renewed attention on content knowledge has also been highlighted for middle school teachers in the United States in light of "No Child Left Behind" (NCLB) legislation requiring all middle school teachers be highly qualified in mathematics or science, which most states have interpreted to mean at least 24 credit hours of pure mathematics or science coursework. The National Science Board's Science and Engineering Indicators 2004 states that:

- 57% of middle school students – studied mathematics with a teacher who did not major or minor in mathematics or a related field.
- 34% of middle school students – received instruction in biology/life sciences from a teacher without a degree in biology, life sciences or a related field.
- 48% of middle school students – received instruction in physical sciences from a teacher without a major or minor in a physical science, engineering or a related field.
- Both high-poverty and high-minority schools had a higher proportion of inexperienced and/or underqualified science teachers than low-poverty and low-minority schools.

How can teachers gain the knowledge of mathematics or science to successfully teach? To address this question, a number of researchers are heavily engaged in specifying the professional knowledge that mathematics and science teachers need and what course work would provide the skills and abilities to develop professional knowledge. The beginning of contemporary efforts to explore teachers' professional knowledge is typically attributed to Lee Shulman and his presidential address at the 1985 annual meeting of the American Educational Research Association, entitled "Those who understand: Knowledge growth in teaching." In the published version of that address, Shulman (1986) suggested that "we distinguish among three categories of content knowledge: (a) subject matter content knowledge, (b) pedagogical content knowledge, and (c) curricular knowledge" (p. 9). Educational scholars have built upon and extended Shulman's seminal work on teachers' professional knowledge, particularly within the domains of subject matter and pedagogical content knowledge (Ball, Thames & Phelps, 2005). This paper provides an account of the difficulty of measuring subject-matter content knowledge and the philosophical and practical problems that occurred in this math and science partnership in the early years as well as possible solutions for the future.

Context: The Rocky Mountain-Middle School Math and Science Partnership

The Rocky Mountain-Middle School Math and Science Partnership is a National Science Foundation-funded, 5-year project that targets middle school teachers and students in seven Denver-area school districts. The project links these school districts with faculty from University of Colorado at Denver and Health Sciences Center's College of Liberal Arts and Sciences and School of Education and Human Development as well as faculty from four other university partners to increase the subject-matter content and pedagogical content knowledge of middle school teachers. In No Child Left Behind language, the project is trying to increase the numbers of "highly qualified" middle school teachers.

The project's primary component is to provide math and science content courses to middle level teachers. Since the project's inception in 2004, fifteen content-based math and science courses have been developed. These courses have been designed to be approximately 80% content and 20% pedagogy and are taught in the summer as 2-3 week institutes. In addition, structured follow-up courses associated with each content course are taught during the academic year. These courses are taught across four Saturdays during a single semester and are focused 80% on pedagogy and 20% on content, the latter consisting of either new or review of summer content based on feedback from the teachers in the content courses or based on Teacher Content Inventory data. In addition, some of these courses are taught in the academic year, usually one math and one science course per semester, that provide teachers the content from the content course and the structured follow-up in one course instead of two. All courses are co-taught by faculty from the College of Liberal Arts and Sciences, School of Education and Human Development, and our K-12 partners.

Table 1. RM-MSMSP Courses Developed

Mathematics Courses	Science Courses
Algebraic Patterns & Functions I & II	Forces and Motion
Math Modeling	Light, Color, and Geometrical Optics
Geometry	Ecology, Biodiversity, & Adaptation
Discovery and Use of the History of Math	Cells, Human Systems, & Heredity
Discrete Math	Earth Processes
Statistics & Probability	Earth Field Experience
Mathematics of Change	Atoms & Properties of Matter
	Interactions of Elements & Compounds

Initially, the project intended to use teacher content inventories (TCI) to ascertain the general science and mathematics knowledge of teachers in the project and develop courses based on that data. However, due to timing of funding and logistics of admitting teachers into the program, assessment could not be given prior to course development thereby driving the course content as intended. Therefore, courses have been co-developed based on the state and national standards and significant input from K-12 partners. Course content was one thing, but the development of appropriate assessments has been difficult, controversial, and complex. Course instructors raised a number of questions, such as:

1. Should we be assessing subject matter only or subject matter and pedagogy; general pedagogical knowledge or pedagogical content knowledge (Grossman, 1990; Ball, 1990); or some other category of "knowledge"?
2. What do we do when a teacher content inventory is valid and reliable but does not address the content of the course or match the standards linked to accountability for our state or school district? Should we be assessing general subject matter knowledge or course specific subject matter knowledge?
3. Should we be assessing for course content if it's a higher level content than what the teacher will be teaching or do we assess for the content that we hope that the teacher will be applying in the middle school classroom?
4. How does the dissonance created by use of multiple tests and adaptations of tests impact evaluation and research findings?

Subsequently, the development team for each course chose the type of teacher content inventory they would use and hence, the project now uses a variety of assessments and we use the term "teacher content inventory" loosely without clear definition. Some instructors use general knowledge assessments developed by national centers including the University of Louisville Center for Research in Mathematics and Science Teacher Development (2004) and Horizon Research (unpublished); others have been adapted by instructors to more closely fit their purposes and course subject matter, while still others were developed by the instructors themselves. Validity and reliability on all of these instruments falls into a variety of categories from extensive but not complete to minimal to none at all.

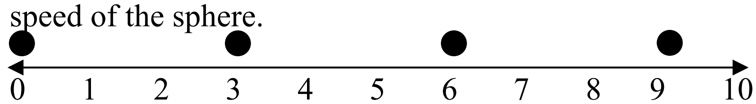
For project teams in both math and science, assessment has been a veritable quandary due to the myriad of ways of thinking about content, what content to teach and what content to assess. In this paper, we will examine the types of items on the different TCI's, the data that have been collected thus far, patterns that we have observed, and implications for the work.

Examining Assessment Questions

Developing the questions for the TCI's is a study in itself as different philosophies and perspectives determined how questions were formulated and what knowledge instructional teams felt was important to assess. For purposes of this paper, we provide a small sample of questions but a very telling sample in terms of how items were constructed and how instructional teams were thinking about teacher content knowledge (see Table 2). These examples illustrate the variety of questions asked: factual knowledge (knowledge about a specific piece of information), conceptual or theoretical knowledge (knowledge about big ideas or models), procedural knowledge (knowledge about science or mathematical processes), knowledge of misconceptions (knowledge of mistaken beliefs), pedagogical content knowledge (the way content is taught within the context of instruction) and self-perception knowledge (teachers' beliefs about their content knowledge).

For example, the Geometry team asked teachers to draw a triangular prism and represent hidden lines with dashes. This question assessed teachers factual knowledge and their ability to pictorially represent a three dimensional figure (see Table 2). On the other hand, the algebra team, that used the Diagnostic Teacher Assessments in Mathematics and Science (DTAMS) (Bush, 2004), posed a question that required teachers to have conceptual understanding as well as strong pedagogical content knowledge to address a question regarding slope (see Table 2). In addition, the History of Math team decided that they were more interested in affect or self-efficacy and created an assessment that measured teachers self-perception of knowledge.

Table 2: Teacher Content Inventory Items

Course	TCI Item	Knowledge Type
Physics I	Flash strobe photographs were taken every second of a sphere moving from left to right. The diagram below shows the location of the sphere when each photograph was taken. The distances are in meters. Find the average speed of the sphere.  A. 9 m/s B. 6 m/s C. 3 m/s D. 0 m/s	Procedural
Physics II	A torpedo shape is an extended source, meaning that it is	Conceptual

	not a point source. A convergent lens is placed near it and an image is made on the screen. Draw what you expect the image to look like.	
Chemistry I	One of your students asks you what the bubbles in boiling water are made of. You reply....	Conceptual (misconception-focused)
Chemistry. II	When an ionic solid dissolves in water, it a. melts into the liquid phase and disperses in the water. b. breaks apart into molecules of the salt which disperse the water. c. breaks apart into positive and negative ions which disperse in the water. d. react with the water to form sodium atoms and chlorine molecules which disperse in the water.	Conceptual (misconception-focused)
Discrete Math	You ask people the month they were born. How many people must you ask to be certain that at least two of them were born in the same month?	Conceptual
Modeling	Suppose you have 1200 feet of fencing to enclose a rectangular corral with one side facing along a river. Assume that the side along the river does not need to be fenced. a. Let *h* be the length of the side of the corral along the river. Find the function that gives the area of	Procedural

	the corral in terms of *h*. b. Plot the area function and determine the value of h that maximizes the area.	
History	I have thorough knowledge of the lives of several mathematicians (Likert scale?). 1 2 3 4 5	None
Geometry	In the space provided on your answer sheet, draw a sketch of a triangular prism. Hidden edges should be drawn with dashed lines.	Factual
Algebra	Your eighth grade students are able to create scatter plots, find the line of best fit (with calculator and estimating), and determine the slope of the line they found. On the other hand, your students seem to have great difficulty explaining what the slope means in terms of the data. Explain how you would help these students understand slope and its applications.	PCK

In discussions with instructional teams the following comments were made that provide some insight into the thinking of particular teams.

> "The Louisville TCI did not match the course being taught. It was a good test but the course material was different than the exam. We did not want students confused or frustrated by questions that the course had

not addressed. Also, we had to make sure the most critical state standards were covered first, that was important to our K-12 partners."

"We are moving more towards constructed response questions and away from multiple choice because we believe conceptual/theoretical knowledge is more important than factual knowledge."

"We used items from a number of valid and reliable instruments even though we know this was dangerous territory. We couldn't use one instrument because one instrument didn't apply to us."

"We were more concerned about how they felt about their learning and the importance of the learning than whether they knew isolated facts about the history of math. We saw what they learned in other assessments and presentations."

Certainly, the individual team purpose of the assessment played a huge role in how instructional teams created and used each teacher content inventory. They were not assessing general teacher content or pedagogical content knowledge, they were really more interested in measuring what teachers gained from their individual course. Data from these assessments, therefore, are a mixed message.

Data and Mixed Messages

This section will describe the data from Years One and Two and provide a general framework for how the project assessments developed, and the questions that were raised by instructors. Development teams worked fairly independently with only one or two group meetings per year. These meetings focused on course content, integration of second language instruction, the learning cycle, co-teaching, and other instructional components we wanted to make sure were modeled in each course. Project leaders suggested that development teams use already developed assessments where they could, but most teams, without an already developed test, created their own.

Year One: 2005

In the first summer of the grant, seven summer academy courses were taught. Teacher content inventories were given to all teachers pre- and post-course. Data from those tests were used to inform the curriculum of the structured follow-up courses; however a second post-test was not given, after the follow up courses. (After examining the content taught in the structured follow-ups, a decision was made in the second summer to add an additional post-test at the conclusion of the structured follow-up. These data will not be available until January 2007 and May 2007.)

In any event, preliminary Year 1 data (Table 3) suggested positive statistical significance on 22 out of 26 subcategories, or overall test scores from a total of six tests given ($p < .05$). The effect sizes ranged from medium to large, indicating the magnitude of the gains. As might be expected, the largest improvements were observed for the courses which used locally developed teacher content inventory instruments (earth science, physics, and geometry) that were directly targeted at course content.

Table 3. Summary TCI Results

Subject	Knowledge Type	Statistical Significance*	Effect Size
Algebra			
(Louisville)	Memorized / Factual Knowledge	No	0.19
	Conceptual Understanding	Yes	0.61
	Reasoning /Problem Solving	No	0.16
	Pedagogical Content Knowledge	Yes	0.59
	Subcategory		
	Patterns, Functions, Relations	Yes	0.33
	Expressions and Formulas	Yes	0.45

	Equations and Inequalities	Yes	0.46
	Overall	Yes	0.45
Biology			
(Louisville)	Declarative	Yes	0.36
	Inquiry	Yes	0.52
	Schematic	No	0.25
	Pedagogical Content	Yes	0.67
	Subcategory		
	Structure / Function	No	0.39
	Internal Regulation	Yes	0.54
	Heredity / Diversity	No	- 0.03
	Interdependence	Yes	0.36
	Overall	Yes	0.45
Chemistry	Overall	N/A**	0.91
(Instructor-created)			
Earth Science	Overall	Yes	3.15
(Instructor-created)			
Geometry	Overall	Yes	1.31
(Instructor-created)			
Physics	Overall	Yes	2.37
(Horizon, Forced Concept Inventory			
Statistics			
(Louisville)			
	Memorized / Factual Knowledge	No	0.22
	Conceptual Understanding	Yes	0.45
	Reasoning /Problem Solving	Yes	0.48
	Pedagogical Content Knowledge	Yes	0.48
	Subcategory		
	Data Analysis	Yes	0.52
	Probability	Yes	0.42
	Overall	Yes	0.53

* All t – tests were one-tailed at significance level $\alpha = 0.05$

** Data not available to compute statistical significance.

As specified previously, these inventories differed in their construction with much variability in reliability and validity depending on the source of development. Specifically, externally-developed Teacher Content Inventories were used in three courses, Algebra, Biology (year 1 only), and Probability and Statistics. These inventories were developed at the University of Louisville Center for Research in Mathematics and Science Teacher Development. The physics course used a combination of questions from an instrument being developed by Horizon Research, Inc and questions created by a variety of previously developed content assessments in physics education (Force Concept Inventory). The remaining courses used locally developed Teacher Content Inventories.

Overall, constructing instruments that match intended course goals and objectives of designers appear to yield very high gains with large effect sizes. However, in this project, these instruments have yet to be determined to be reliable or valid. The external instruments with some established reliability and validity did not align well with the course content and goals..

A more in-depth comparison reveals huge differences not only in the results for each course but also in what we can learn about teachers' overall knowledge. Table 4 is an example of the Algebra teacher content inventory created by researchers at Louisville University and illustrates the complexity of the external teacher content inventory and the multiple sub-measures involved.

Table 4: Algebra

			Pretest		Posttest	
	N	Max Possible Score	Mean	S.D.	Mean	S.D.
Knowledge Type						
Memorized/Factual Knowledge	25	10	**6.84**	2.08	**7.2**	1.71
Conceptual Understanding	25	10	**5.68**	2.58	**7.12**	2.09
Reasoning/Problem Solving	25	10	**5.72**	2.48	**6.08**	2.08
Pedagogical Content Knowledge	25	10	**2.52**	2.16	**4.0**	2.83
Subcategory						
Patterns, Functions, Relations	25	20	**11.48**	4.21	**12.84**	4.03
Expressions and Formulas	25	9	**4.24**	2.19	**5.12**	1.69
Equations and Inequalities	25	11	**5.04**	3.02	**6.44**	3.02
Overall Score	**25**	**40**	**20.76**	**8.54**	**24.4**	**7.63**

When scores on the pretest and posttest were compared using t tests, the posttest yielded a significantly higher score than did the pretest. The mean gain was 3.64 points. The group of participants who completed the algebra course showed significant improvement in overall learning, as measured by the TCI pre- and post- test results. They also made significant gains in all the subcategories tested as well as in two out of four knowledge types (conceptual understanding and pedagogical content knowledge). The fairly large effect sizes associated with the conceptual understanding and the pedagogical content knowledge are indicative of substantial gains made in these two areas.

In comparison, Table 5 illustrates the results from the geometry course where course instructors designed their own assessment tool and only an overall score was determined as there were no subcategories.

Table 5. Geometry

			Pretest		Posttest	
	N	Max Possible Score	Mean	S.D.	Mean	S.D.
Overall Score	**23**	**34**	**17.39**	**6.23**	**25.09**	**5.49**

The posttest yielded a significantly higher score than did the pretest. The mean gain was 7.7 points and the effect size was 1.31. The participants who attended the geometry/measurement course showed significant improvement in overall learning. The effect size was very large reflecting the very substantial gain made by the participants.

Although both courses show statistically significant gains, the geometry course showed higher gains. When the results of these exams are examined closely and in relation to the course content that was taught, results are difficult to compare due to the complexity of the nature and reasons for the particular assessment used. In the algebra assessment we learn more about different “knowledges”, but it is not necessarily reflective of the course content or the context of the teachers taking the course. In the geometry course, we learn more about whether teachers learned the specific content taught in the course, but nothing about the different types of knowledge.

Year Two: 2006

In Year Two, some of the assessments changed again as instructors felt that they wanted more information about whether the content in their specific course was making a difference for teachers. Some teams decided not to use the Louisville DTAMS in lieu of their own test, and some with the onset of additional courses in the second year (7 to 15) the problem with assessment was exacerbated.

Table 6 provides a description of each assessment used in the project, where it originated, progress of validity and reliability, and the types of questions (content or pedagogical content) included on the assessment.

Table 6. Assessment Description

Course	Originated	Validity and Reliability Testing	Content Qs	Pedagogical Content Qs
Physics I & II	Horizon	In progress	Yes	Yes
	Instructor	Not conducted	Yes	No
Statistics and Probability	Louisville	In progress	Yes	Yes
Biology I	Louisville in Year 1; Instructors in Year 2	In progress on both	Yes	Yes
Ecology I	Instructors	Beginning work		
Chemistry I	Instructors	Not conducted	Yes	Yes
Chemistry II	Instructors	Not conducted	Yes	Yes
Discrete Mathematics	Instructors	Not conducted	Yes	No
Mathematical Modeling	Instructors	Not conducted	Yes	No
History of Mathematics	Instructors	Not conducted	No (This assessment was designed to assess self-efficacy of content.)	No
Geometry	Instructors	Not conducted	Yes	No
Algebraic Ideas	Louisville	In progress	Yes	Yes
Mathematics of Change	Instructors	In progress	Yes (Algebra)	Yes
Statistics	Louisville	In progress	Yes	Yes
Geology	Instructors	Not conducted	Yes	No
Earth Processes II	Instructors	Not conducted	Yes	No

Table 7 provides statistical data for each course. Again, T-test for matched groups (paired t-tests) were conducted to determine whether knowledge and skills related to geometry/measurement increased, decreased, or stayed the same from the pre-test to the post-test among students. All tests were conducted at the 0.05 significance level. Since it is hypothesized that participants' knowledge and skills will improve after enrolment in the course, one-tailed tests were performed in all cases. The average normalized gain was also calculated for this group. The normalized gain index is calculated by dividing the difference in posttest score and pretest score by the difference between 100% and the pretest score. The effect size (*ES*), or Cohen's d, was used as a measure of gain from pretest to posttest.

Table 7. Teacher Content Inventory 2006 – Summary of Results

Subject	Max Possible Score	Average Gain	Percent Gain	Statistical Significance*	Normalized Gain	Effect Size
Biology I	40	9.91	24.78	Yes	0.48	1.41
Ecology I	44	8.52	19.36	Yes	0.35	1.25
Chemistry I	N/A	N/A	N/A	N/A	N/A	N/A
Chemistry II	N/A	N/A	N/A	N/A	N/A	N/A
Discrete Mathematics	12	3.38	28.17	Yes	0.49	1.69
Mathematical Modeling	139	36.74	26.43	Yes	0.49	1.68
History of Mathematics	45	16.98	37.73	Yes	0.74	3.35
Geometry	100	20.72	20.72	Yes	0.45	1.18
Algebraic Ideas	N/A	N/A	N/A	N/A	N/A	N/A
Mathematics of Change	N/A	N/A	N/A	N/A	N/A	N/A
Statistics	40	3.57	8.93	Yes	0.42	0.70
Geology	75	19.1	25.47	Yes	0.58	1.89
Earth Processes II	43	26.84	62.42	Yes	0.91	3.80
Physics I	25	5.56	22.24	Yes	0.43	1.50
Physics II	20	6.8	34	Yes	0.66	2.23

* all significance at the 0.001 level.

Not surprisingly, we found significant gains on each assessment that has been analyzed so far. But again, making overall assumptions or stating implications of the work is difficult.

Lessons Learned

Assessing teacher content knowledge presents a conundrum that is not easily rectified and highlights the question of the specific knowledge needed to be addressed in mathematics and science courses specifically designed for middle school teachers. On the issue of what types of knowledge to assess, the majority of instructional teams believe that we should assess content knowledge, pedagogical content knowledge, and instructional content knowledge, but not at one time or within the jurisdiction of one course. Over the duration of the project, however, other assessments are being utilized to assess these types of knowledge since we recognize that the highest quality of teaching requires all of them.

The second issue concerns validity/reliability vs. course criterion referenced assessment. In almost all cases, except Algebra and the Math of Change course, instructors find more value in creating their own assessments rather than using valid and reliable assessments created by research and development centers. They have opted for depth rather than breadth in content especially in the sciences. University science faculty participating in the project are still concerned that the "depth" of content is not deep enough nor that the content is presented at a significantly high enough level (the sentence seems awkward to me). Some of this concern is attributable to the fact that we offer these courses for graduate credit and count them towards a Masters degree in the School of Education and Human Development. K-12 and School of Education faculty are concerned about whether course content is meeting State standards and whether the content is the "right" content for middle school teachers.

In addition, one of the fundamental philosophical tenets of project participants is that by providing teachers a higher level of understanding in content, beyond what they would be expected to teach, will help them provide better instruction in the middle school content. For example, in the Math of Change course, the instructional team has decided to assess Algebraic knowledge instead of knowledge in Calculus (the real content of the course). They anticipate that by knowing and understanding more about higher mathematics, teachers will be able to make better connections for their students in Algebra.

There is also the issue of misconceptions. Instructional teams worried about the misconceptions held by teachers and how those misconceptions were seeping into their classrooms. Several of the courses (biology & chemistry) used pre/posttest questions that aimed to probe the misconceptions held by teachers and/or their students (Kind, 2004). Multiple choice versions used what Philip Sadler calls "attractive" distractors in that they contain common misconceptions about scientific concepts (1998). We also asked open-ended questions that probed common scientific misconceptions. (Examples of each are found in Table 2 under Chemistry I and II). Course instructors were able to directly identify misconceptions held by teachers and address them through the course content. For example, fully half of the middle school teachers did not understand that the bubbles in boiling water contained water vapor, rather responding that they contained air, hydrogen and/or oxygen gas, or nothing. The course content included an extensive coverage of phase changes and posttest results revealed complete comprehension of this aspect of phase changes.

And finally, how does the dissonance created by use of multiple tests and adaptation of tests impact evaluation and research findings? There is no question, that there is great tension between the practicality of meeting project goals and the research into the effectiveness of our approach when it comes to teacher content inventories. It may only be when all the data are collected and we have a grander vision of the outcomes of the project, that we'll know if the latitude that we've given instructors and the care that we've taken to create an atmosphere for teachers where they do not get discouraged and want to continue their content learning was the right thing to do.

As in every discipline (content and pedagogy), we are faced with the impossibility of teachers knowing everything that the collective "we" knows, thus decisions about content knowledge, content pedagogical knowledge, and instructional knowledge need to made carefully, based on the project's desired outcomes and the context of the participant teaching population. To this end, we need to remember that these assessments are only one measure of the total quality of any one teacher. We must look at a "package" of assessments to view the whole teacher and continue to probe what works best, how teachers learn best, and how this new learning impacts students.

Implications for Future Work

As the management team and instructional teams ponder the data and think about the next three years of this project, it is clear that we need some consistency in assessment if we are to provide any clear evidence of increased teacher content knowledge. We will, however, need to carefully balance our research and evaluation needs with the needs of participating teachers. The first question we will need to address is what knowledge(s) do we want to measure across all participants in mathematics and science. We have teachers taking multiple courses in both math and science, (a few in both content areas). We would like to know if how each course impacts general knowledge in other sub-disciplines as well as the one(s) the teacher is participating in. In mathematics, we are going to put together a general test using questions already developed. In science, we will have to do the same for the same reason, but we know that one general knowledge test in science is much more difficult.

On a different level, the course-referenced tests that have been created for each course are useful to instructors as they continue to work on and improve course content, but again we believe that there should be some consistency in the types of questions that are asked and the types of knowledge that assess.

Maintaining balance in the project, not overwhelming teachers with assessment, ensuring that the courses are high quality for sustainability are key to the project's success. Prior to next summer's courses, we have a lot of work to do but feel that now we have some practical application behind us and that philosophically we are moving together in the right direction.

References

Ball, D.L., Hill, H.C. & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? American Educator, 14-46.

Ball, D.L., Thames, M.H. & Phelps, G. (April, 2005). Articulating domains of mathematical knowledge for teaching. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec.

Ball, D.L. (1990). The mathematical understandings that prospective teachers bring to teacher education. The Elementary School Journal, 90, 449-466.

Bush, W. (2004). Diagnostic teacher assessments in mathematics and science. University of Louisville Center for Research in Mathematics and Science Teacher Development.

Horizon Research, Inc. (Unpublished) . http://www.horizon-research.com/

Kind, V. (2004). Beyond Appearances: Students' Misconceptions about Basic Chemical Ideas, 2nd Ed., Chemical Soc.: London.

Loucks-Horsley, S., Hewson, P., Love, N., Stiles, K., (1998). Designing professional development for teachers of science and mathematics. Thousand Oaks, CA: Corwin.

Grossman, P (1990). The making of a teacher: Teacher knowledge and teacher education. New York: Teachers College Press.

Jacobs, J., Hiebert, J., Givvin, K.B., Hollingsworth, H., Garnier, H., & Wearne, D. (2006). Does eighth-grade teaching in the United States align with the NCTM

Standards? Results from the TIMSS 1995 and 1999 Video Studies. Journal for Research in Mathematics Education, 37(1), 5-32.

Mewborn, D.S. (2003). Teaching, teachers' knowledge, and their professional development. In J. Kilpatrick, G. Martin, and D. Schifter (Eds.), A research companion to Principles and standards for school mathematics (pp. 45-52). Reston, VA: NCTM.

National Committee on Science Education Standards and Assessment, National Research Council (1996). National science standards. Washington, DC: National Academies Press

National Council of Teachers of Mathematics. *Principles and Standards for School Mathematics*. Reston, Va.: National Council of Teachers of Mathematics, 2000.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

National Science Board (2004). Science and Engineering Indicators 2004. Washington, DC: National Science Foundation. Retrieved July 29, 2004, from http://www.nsf.gov/sbe/srs/seind04/start.htm

Sadler, P.M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *J. Res. Sci. Teach.* **35**:265-296.

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