Strengthening and Assessing Teachers' Physics Content Knowledge

Purpose

The No Child Left Behind (NCLB) Act of 2001 required states to ensure that all teachers of core academic subjects in public schools be "highly qualified" by the end of the 2005-06 school year. However, a shortage of certified science teachers in the United States has led many states to develop alternative routes to teaching, including attempts to convert teachers from content areas of over-supply into science teachers. Such teachers may not have the content background to be highly qualified for teaching science, generating the need for programs and interventions to enhance science content knowledge for groups of teachers. High quality direct assessments of teachers' content knowledge and pedagogical content knowledge are essential tools the science teacher education field needs in order to monitor and assess efforts to strengthen this important characteristic of science teacher knowledge.

Framework: Two Goals of Paper

This paper has two foci: one focusing on the profile of scores available from the Diagnostic Teacher Assessments in Mathematics and Science (DTAMS) assessments due to the integral structure of the assessments, and the other focusing on how this structure enhanced interpretability of course impact in a study of teacher physics content knowledge development. The goal of the focus on the DTAMS score profiles was to provide insight into the underlying rationale and emphases of these assessments (three assessments available in physical science, life science, and earth/space science). Highlighting the score profiles returned by the assessments enables those who may wish to consider using these assessments of teacher content knowledge for their projects to more easily evaluate the alignment of their project with the design of the assessments. The second goal, reporting on a study of course impact on teachers' physics content knowledge, reported the study's outcomes in the framework of the information available from the assessments. This framework provided a reasonably fine-grained analysis that may be useful for interpreting nuances of the impact of the course.

Methods and Data Sources: Structure of DTAMS Assessments and Subscales

An extensive, multi-year development process resulted in the generation of science teacher content knowledge assessments (DTAMS) that are grounded in several sources of validity evidence and that have reported solid internal consistency measures of Cronbach alpha > .81 based on a sample of approximately 1800 teachers across multiple states (Author, 2007). The validity evidence includes: item development by teams of science educators, scientists, and science teachers; external review of all items by panels of a couple dozen experts; field testing and revision with several hundred teachers in multiple states (Author, under review).

The internal structure of the assessments was intentionally constructed to provide results on two independent dimensions: the breadth and the depth of content knowledge. Breadth of content knowledge was established by grounding the content through extensive synthesis across a spectrum of eight national and international documents, including both curriculum standards documents as well as national and international student assessment frameworks. The standards documents included: the *National Science Education Standards* (NSES) (National Research Council, 1996); *Benchmarks for Science Literacy* (American Association for the Advancement of Science, [AAAS], 1993); and Mid-continent Research for Education and Learning's [McREL's] science portion of *Content knowledge: A compendium of standards and benchmarks for K-12 education* (Kendall & Marzano, 2004). Student assessment frameworks used to synthesize content recommendations included: the *National Assessment for Educational Progress* (NAEP) (National Assessment Governing Board [NAGB], 2004); and *Trends in International Mathematics and Science Study* (TIMSS) (Mullis et al., 2003).

In addition to student standards, teacher content knowledge recommendations were reviewed. The teacher content knowledge standards analyzed were from the National Science Teachers Association (NSTA, 2003) and the Interstate New Teacher Assessment and Support Consortium (INTASC) Science Standards Drafting Committee (INTASC, 2002). Lastly, *The Praxis Series: Professional Assessments for Beginning Teachers* (2004) was examined for teacher content knowledge recommendations. See Author (2008) for a detailed description of the synthesis process across these eight documents.

Depth of content knowledge was structured on a taxonomy built on the work of Li and Shavelson (2001), Stiggins (1997), and Shulman (1986). This process resulted in four types of knowledge: (a) declarative (rote memorization, factual knowledge), (b) inquiry (skills of doing science, process skills), (c) schematic (understanding of underlying concepts, principles, theories and the connections between ideas), and (d) pedagogical content knowledge.

Methods and Data Sources: Study Utilizing Information from these Assessments

The content knowledge impact of a semester-long "Physics for Teachers" course was investigated using a pre-post design, with two equivalent versions of DTAMS Physical Science used as the measurement. The philosophy undergirding the course was an emphasis on deep conceptual understanding of core physics concepts rather than mathematical formulation, and included a strong secondary focus on the inclusion of pedagogical techniques that target the adult learners but are in many cases directly adaptable for the teachers' own K-12 classrooms. One technique systematically employed was to investigate physics concepts in the context of using everyday applications as starting points. For example, one piece of the course studied how a refrigerator worked to introduce and then explore physical science topics related to thermodynamics and gas laws.

The sample for this study was 43 teachers (5 elementary certified, 21 middle school certified, and 17 high school certified). Of these participants, 9 held certification in a physical science (chemistry or physics) whereas the majority (32) held certification either in general science or in a non-physical science domain such as biology. With the exception of two outlier teachers each with 20 years experience, this group of teachers was primarily in the early parts of their careers with a mean of 2.7 years experience (standard deviation = 2.6).

Results and Brief Discussion

Results in Table 1 are reported by total score, and subsequently by the set of subscores comprising each of the independent dimensions of depth and breadth of content knowledge. In addition to reporting results of testing for statistically significant growth from pretest to posttest (with the semester course serving as the intervention), Table 1 included effect sizes (partial eta-squared) to aid in the interpretation of results.

Table 1
Growth in Physical Science Scores (n=43)

Subscale (pts. possible)	Pre	Post	Partial eta-squared effect size
Total (35)	17.84 (4.74)	27.77 (5.33) ***	.694
	Depth of Knowledge Subscales		
Declarative (5)	3.67 (1.06)	4.28 (1.03) **	.151
Inquiry (5)	3.23 (0.97)	3.07 (1.30)	.012
Schematic (15)	7.88 (2.34)	12.09 (2.46) ***	.627
Pedagogical (10)	3.05 (2.12)	8.33 (1.71) ***	.814
	Breadth of Knowledge Subscales		
Matter (6)	4.16 (1.15)	4.86 (1.08) **	.197
Motion & Forces (6)	3.86 (1.04)	4.70 (1.21) ***	.243
Energy (13)	6.77 (2.07)	9.88 (2.27) ***	.527
STS (5)	2.42 (1.26)	3.79 (1.10) ***	.407

Note. Data are mean (standard deviation) of the various subscales, but subscales cannot be directly compared because the total points possible varies across subscales.
*p<.05. **p<.01. *** p<.001.

The total test score change showed a statistically significant growth, and the effect size showed that growth to be quite substantial. However, if the analysis had terminated at that point, additional important details would have been missed. As shown in Table 1, the growth was not evenly distributed across the depth of knowledge dimension, with the bulk of the growth (as indicated by effect sizes) concentrated in the schematic knowledge (the deep, conceptual understanding) and the pedagogical knowledge. Although declarative knowledge also resulted in a statistically significant growth which might have led to interpretations including that subscore in the list of impacts, the relatively small effect size compared to the two largest suggests that this particular subscore growth was not as substantial as the other two. This distinction illustrates the need to consider effect sizes along with significance when interpreting results. Because the course design intentionally focused on schematic knowledge (deep conceptual understanding) as well as incorporating pedagogically powerful and effective techniques to deliver this content, the substantial growth in those two subscores confirms the successful impact the course had in these areas.

Similarly for the breadth of knowledge subscore spectrum, growth was not uniformly distributed. Although the tests of statistical significance suggest that all four might have been interpreted to contribute to the overall growth, the differences in effect sizes suggested that the content areas of energy and STS contributed disproportionately to the growth. Because the course emphasis was on unifying concepts (such as energy considerations) and contextualized around everyday applications (STS), the substantially stronger growth in these two subscores is confirmation that the course was reasonably successful in those two areas.

These results reinforce that the measurement instrument was sensitive to the intended emphases of this course. This, combined with the validity and reliability evidence reported elsewhere for DTAMS (Author, under review), suggested that this assessment may provide a useful tool for measurement of teacher science content knowledge.

Scholarly Significance

The structure of the DTAMS assessments provided detailed measures across breadth and depth spectra of content knowledge, enhancing their potential usefulness as diagnostic measures of teacher content knowledge. The study investigating the impact of a semester-long course intervention for a sample of 43 science teachers serves as an example of a potential use and interpretive lens that may be applied to results from this measure. The combination of a more thorough understanding of the assessment and the scores it produces, together with an example of a study utilizing and interpreting these scores, will enable researchers to more confidently evaluate the potential usefulness of these teacher content knowledge assessments in their own work.

Valid and reliable assessments of teacher science content knowledge provide access to direct measurement of a crucial variable of interest to educational researchers, professional development providers, and teachers. The DTAMS assessments (available for physical science, life science, and earth/space science) could be used to determine the impact of workshops, courses, or other experiences on teachers' knowledge. These valid and reliable measures of teacher knowledge are needed to compare teacher knowledge in a variety of research scenarios. Teachers will be able to use the diagnostic assessments to determine their strengths and weaknesses, and subsequently, to make informed choices about further coursework or other professional development that may be most beneficial for them.

References

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy.* New York: Oxford University Press, Inc.

Author (2007). Paper presented at AERA international conference.

Author (2008). *Journal of Science Teacher Education*.

Author (under review). *Journal of Research in Science Teaching.*

Interstate New Teacher Assessment and Support Consortium (INTASC) Science Standards Drafting Committee. (2002). *Model standards in science for beginning teacher licensing and development: A resource for state dialogue*. Council of Chief State School Officers. Retrieved January 14, 2004 from http://www.ccsso.org/content/pdfs/ScienceStandards.pdf

Kendall, J. S., & Marzano, R. J. (2004). *Content knowledge: A compendium of standards and benchmarks for K-12 education*. Aurora, CO: Mid-continent Research for Education and Learning. Online database: http://www.mcrel.org/standards-benchmarks/

Li, M., & Shavelson, R. J. (2001). *Examining the links between science achievement and assessment.* Paper presented at the AERA Annual Meeting, Seattle, WA.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., et al. (2003). *TIMSS Assessment Frameworks and Specifications 2003* (2nd ed.). Retrieved January 14, 2004 from http://isc.bc.edu/timss2003i/PDF/t03_AF_sci.pdf

National Assessment Governing Board. (2004). *Science framework for the 2005 National Assessment for Educational Progress.* Washington D.C.: Author.

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

National Science Teachers Association (NSTA). (2003). *Standards for science teacher preparation. NSTA.* Retrieved January 14, 2004 from http://www.nsta.org/main/pdfs/NSTAstandards2003.pdf

Praxis Series: Professional Assessments for Beginning Teachers. (2004). *General Science Content Knowledge Part 2 (0432)*. Educational Testing Service (ETS). Retrieved January 14, 2004 from http://www.ets.org/Media/Tests/PRAXIS/pdf/0432.pdf

Shulman, L. (1986). Those who understand, knowledge growth in teaching. *Educational Researcher, 15*(2), 4−14.

Stiggins, R. J. (1997). Performance assessment of skill and product outcomes. In K. M. Davis (Ed.), *Student-centered classroom assessment* (2nd ed., pp. 261−302). Columbus, OH: Prentice-Hall, Inc.