4-14-2009

Can Science Teachers' Strategic Knowledge be Conceptualized as a Learning Progression?

Robert M. Talbot III
University of Colorado Denver, robert.talbot@ucdenver.edu

Derek Briggs

Valerie K. Otero

Recommended Citation

Talbot, R. M., Briggs, D. C., & Otero, V. K. (2009, April). Can science teachers' strategic knowledge be conceptualized as a learning progression? Symposium presentation at the annual meeting of the American Educational Research Association Annual Meeting, San Diego, CA.

Can Science Teachers' Strategic Knowledge be Conceptualized as a Learning Progression?

Paper Presented at the annual meeting of the American Educational Research Association

April 14, 2009

Robert M. Talbot III

Derek C. Briggs

Valerie K. Otero

Introduction

As the organization of this special symposium on *Learning Progressions to Describe Teacher Development* makes clear, learning progressions have become a hot topic. This is true in the context of educational curriculum development and assessment in general, and science education in particular. One of the most authoritative and widely cited definitions of learning progressions to date comes from the 2006 National Research Council report *Systems for State Science Assessment*: “Learning progressions are descriptions of the successively more sophisticated ways of thinking about an idea that follow one another as students learn” (Wilson & Bertenthal, 2006, p. 48) Similar sorts of definitions can be found in *Taking Science to School* (National Research Council Committee on Science Learning Kindergarten Through Eighth Grade., Duschl, Schweingruber, & Shouse, 2007), and by the National Assessment Governing Board (NAGB) in the framework and specifications for the 2009 National Assessment of Educational Progress (NAEP) science test (National Assessment Governing Board, WestEd, & Council of Chief State School Officers, 2007a, 2007b).

In this paper we introduce a survey instrument we have recently developed and tested, the *Flexible Application of Student-Centered Instruction* (FASCI). We provide some empirical results from a pilot test of the instrument and speculate about the assumptions that would need to be met before individual differences in FASCI scores can be interpreted as manifestations of an underlying learning progression. The FASCI instrument consists of a series of classroom scenario-based items (e.g., vignettes) used to elicit scorable information about the way teachers conceptualize, apply, and revise instructional strategies in a classroom learning context. We refer to this as a teacher’s “strategic knowledge.” An underlying hypothesis being operationalized in the way that FASCI item responses are scored is that certain ways of conceptualizing the application of instructional strategies represent qualitatively distinct levels of strategic knowledge. A deeper, implicit assumption for FASCI scores to be interpreted relative to a learning progression is that the development of this strategic knowledge is sensitive to both natural maturation as a teacher grows, and to the potential effects of specific professional training programs. The notion of qualitatively measurable differences in "levels of sophistication" is fundamental to conceptions of learning progressions as models of the ways learners develop their understanding of scientific concepts and processes. In the context of teacher education as well as in practice, teachers are themselves learners in the sense that their knowledge of teaching and learning is continually changing and becoming more sophisticated. Therefore, if the notion of learning progression has potential in the context of student assessment, it should have equal potential as a means for providing insights about the ways that teachers develop their understanding of how to effectively *teach* scientific concepts and processes.

There are three purposes for this paper. The first is to present and discuss the foundations of the FASCI construct. The second is to present some preliminary evidence which shows that respondents can be distinguished from one another on the FASCI construct both qualitatively and quantitatively. The third purpose is to illustrate more generally issues that arise when one takes a measurement-based approach to the conceptualization and development of a learning progression. Under this approach the development of a survey instrument serves as a vehicle for explicating the theoretical and empirical basis that supports (or fails to support) a hypothesized learning progression.

The Motivation and Purposes Underlying the FASCI Instrument

Our initial impetus to develop an instrument to quantify some aspect of the knowledge and skills of teachers was motivated by our work in the Science, Technology, Engineering, Mathematics (STEM) Teacher Preparation project (Otero, 2006). The NSF-funded[1] STEM Teacher Preparation (STEM-TP) program initially began in Fall of 2003 as an multi-disciplinary collaboration of faculty from the College of Arts & Sciences and the School of Education at the at our university, as well as local K-12 teachers. The purpose of the STEM-TP program is to integrate efforts to transform large-enrollment, undergraduate math and science courses to be more student-centered and interactive with efforts to recruit and prepare undergraduate math and science majors to become K-12 teachers.

A key operating premise behind the STEM–TP program is that the best way to develop a qualified pool of undergraduate science and math majors with a possible interest in future teaching careers is to place them in situations where they are directly involved in the teaching and learning process as undergraduate *Learning Assistants* (LAs). LAs receive a stipend for working 10 hours per week in three aspects of the transformation of these large-enrollment courses. First, they lead learning teams of 6-20 students that meet weekly to work on collaborative and inquiry-based activities, such as the Tutorials for Introductory Physics (McDermott & Shaffer, 2002). Second, the team of LAs from each course meets weekly with the faculty instructor of the transformed class in which they work to plan for the upcoming week, to reflect on the previous week, and to examine student achievement in these courses. Finally, LAs from all science departments attend a specialized course, *Mathematics and Science Education* that complements their LA-teaching experiences. In this course, LAs reflect on their own

---

[1] National Science Foundation Grant DUE-0302134.

teaching practices, evaluate the transformations of courses, and investigate relevant educational literature. This course strongly emphasizes the role of student prior knowledge in instruction and LAs are taught to value students' thinking and learn to consider their students' incomplete and experience-based ideas as resources for further instruction and learning. That is, they are taught the formative assessment process (Atkin, Black, & Coffey, 2001) in which teachers elicit and respond to their students' prior-knowledge.

Approximately 15% of students who participate as LAs enter teacher certification programs and become K-12 teachers. LAs who decide to become teachers are eligible for NSF-funded Noyce Teaching Fellowships[2] of up to $10,000 per year. Noyce Teaching Fellows can become Lead LAs who mentor novice LAs, participate in the curriculum development and the development of course educational technology, or work with mathematics, science, and education faculty conducting educational research. Past, present, and future Noyce Fellows meet monthly to discuss current classroom successes and obstacles.

Through the collective experiences of teaching as an LA, instructional planning with a science faculty member, and reflecting on their teaching, it is anticipated that LAs integrate understanding of content, pedagogy, and practice, or what Shulman referred to as *pedagogical content knowledge* (PCK; Shulman, 1986, 1987) which is thought to be a critical characteristic of effective teachers. Putnam and Borko (2000) have demonstrated that instruction on pedagogy is most effective when it is situated in practice, that is, when teachers have the opportunity to try out and revise pedagogical techniques by implementing them with real students. The development of the FASCI instrument was motivated by our desire to measure the effect of the "LA/Noyce Treatment" on participants' knowledge of teaching. We sought to measure the type

---

[2] National Science Foundation Grant DUE-0434144

of teacher knowledge that would cause two different teachers to make different decisions in a similar classroom context. One of the foundations of this type of knowledge comes from the construct of PCK. The link between PCK and the FASCI construct are further explicated in the next section.

Theoretical and Empirical Foundations of the FASCI Construct

The construct of PCK is widely used and referenced in studies on teacher knowledge. As will be shown, the FASCI construct is linked to the construct of PCK and to models which describe the development of PCK. Therefore, we begin this section with a presentation of Shulman’s definition of PCK, and follow that with a discussion of subsequent clarifications and re-definitions of this construct by other researchers. Within this section we also summarize some empirical studies which have invoked this construct. We identify two common aspects of these conceptualizations of PCK, as well as a common context in which these characterizations are made. We then turn to a discussion of models which may help to explain the development or growth of PCK. Specifically, we present Shulman’s model of pedagogical reasoning (Shulman, 1987) and some of the literature on adaptive expertise (Hatano & Inagaki, 1986). Next, we discuss the FASCI construct in order to link the two dimensions which compose it back to the discussion on PCK, and in order to further justify the levels on the construct.

Definitions and uses of PCK

Shulman’s introduction of PCK has proven to be very popular as a construct which describes teacher knowledge. For example, the *Standards for Science Teacher Preparation* (National Science Teachers Association, 2003) is founded in part on this construct. Shulman’s research program “Knowledge Growth in Teaching” focused on the knowledge development of secondary teachers from a variety of disciplines: English, Biology, Mathematics, and Social

Studies. The participants (number unknown) all had bachelor's degrees in their subject area, were part of a cohort of prospective teachers at Stanford, and were followed during their teacher preparation program and into their beginning teacher practice. Each participant was part of the study for 1-2 years. The researchers attempted to "trace the intellectual biographies" of these participants by conducting interviews with them, observing their teaching, and collecting information on the teacher education program in which these teachers were educated. In order to assess participants' content knowledge, the researchers used the intellectual biographies to define the "sources" of the participants' comprehension of subject matter (1986, p. 8). Data collection focused on what Shulman calls "strategic research sites and key events" which were critical to the knowledge development of these novice teachers. One example of a key event is an instance when a teacher needs to teach a subject which they have never learned. This event gives rise to questions about how teachers learn the subjects that they need to teach and how they prepare to teach those topics. Based on what was discovered from this research program, Shulman defined the construct of PCK as "that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding…[and] the ways of representing and formulating the subject matter that make it comprehensible to others" (1986, p. 8-9).

Subsequently, researchers have sought to refine or re-define the construct of PCK. In order to address what they perceived as inconsistencies between Shulman's model of PCK and constructivist perspectives on learning, Cochran, DeRuiter, and King (1993) present an expanded definition of PCK. They contend that content knowledge and PCK are unable to be distinguished from one another within a constructivist framework, and lean towards the idea that all teacher knowledge is pedagogical knowledge in some sense. The choice of the term *Knowing* rather than

Knowledge is meant to convey the idea that teaching involves helping students to generate there own conceptual understanding (p. 266). Their construct of PCKg is defined as "a teacher's integrated understanding of four components of pedagogy, subject matter knowledge, student characteristics, and the environmental context of learning" (p. 266). Cochran et al. believe that Shulman did not emphasize enough the importance of these last two aspects (understanding of student characteristics and of the environmental context of learning), but instead he discussed them within his presentation of *transformation* which "veiled their importance." As presented in Table 1 below, others (e.g., Hashweh, 1987; van Driel, Verloop, & de Vos, 1998) have also focused on the importance of understanding students *within* Shulman's definition of PCK (i.e. without creating a new model or term for a similar construct).

In contrast to the re-definition of PCK presented by Cochran et al. is a definition of science teacher PCK presented by Magnusson, Krajcik, and Borko (1999). These authors focus on the idea of *transformation* of several types of knowledge. They define the construct as:

> a teacher's understanding of how to help students understand specific subject matter. It includes knowledge of how particular subject matter topics, problems, and issues can be organized, represented, and adapted to the diverse interests and abilities or learners, and then presented for instruction. (p. 96)

Based on Shulman's initial definition of PCK and on the works of Grossman (1990) and Tamir (1988), Magnusson et al. present a definition of PCK specifically for science teaching which consists of five components: (a) orientations toward science teaching, (b) knowledge and beliefs about science curriculum, (c) knowledge and beliefs about students' understanding of specific science topics, (d) knowledge and beliefs about assessment in science, and (e) knowledge and beliefs about instructional strategies for teaching science. Of these components, "orientations toward science teaching" is seen by Magnusson et al. as the overarching conception which in turn shapes the other components related to a teacher's knowledge and beliefs about various aspects of science teaching. Each of these components is further described and delineated in the authors' description of this definition. The authors state that in order for science teachers to be effective, they "need to develop knowledge with respect to <u>all</u> of these aspects of PCK" (p. 115, emphasis in original).

Magnusson et al. defend their definition of PCK as one that is able to "guide the focus and design of pre-service and in-service teacher education programs" (p. 116) for two reasons: (a) "its conceptualization as knowledge that results from a transformation of other domains of knowledge signals that it is more than the sum of its parts," and (b) "because this knowledge is conceptualized as being constructed through the processes of planning, reflection, and teaching specific subject matter, it represents knowledge that is 'uniquely the province of teachers, their own special form of professional understanding (Shulman, 1987, p. 8)'" (Magnusson, et al., 1999, p. 116).

In identifying empirical studies which invoke the construct of PCK in the study of teacher knowledge, we first searched for those which had referenced the work of Shulman and/or PCK. Within those studies, we selected those that focused on science teachers and also appeared in peer-reviewed journals, and read the citations and abstracts for these papers. We read those that we felt were the most promising (i.e. the ones that did not merely mention PCK, but invoked the construct in order to frame their research and/or to explain their findings), and also collected and read other papers which were cited by these but which did not come up in our initial search. A summary of these studies is shown in Table 1.

INSERT TABLE 1 HERE

Two common characteristics of PCK fall out of the definitions and empirical studies discussed above: 1) teachers’ representations of subject matter and/or strategies for teaching the subject matter, and 2) teachers’ knowledge of students’ understanding of the subject matter. Inherent in both of these common characteristics are pedagogical knowledge *and* subject matter knowledge, both situated within contextual knowledge. As well, these characteristics often come from research which focuses on the context of a “key event” (in the case of Shulman and Sanders et al.) or a “critical incident” (in the case of Hashweh), which is similar to the context provided in the FASCI scenario-based items.

We have discussed definitions and characteristics of PCK as presented in both theoretical and empirical work. Now we turn to discussing models for the development of growth of PCK so that we are able to link those characteristics back to the FASCI construct in order to further describe levels of sophistication on that construct, and to support our hypothesized learning progression for science teacher strategic knowledge.

Models of growth and development of PCK

Shulman (1987) presented a model of pedagogical reasoning which serves to explain how teachers might progress in their development of PCK. This model also provides a basis for thinking about the judgments and actions made by teachers. Shulman’s model of pedagogical reasoning (shown graphically in Figure 1) posits stages which culminate in new comprehension of subject matter and pedagogy related to that subject matter. Each of these stages involves both knowledge of representations of subject matter and knowledge of students, as well as the notion of adaptation. These stages can be linked to each dimension of the FASCI construct, as will be discussed further below.

INSERT FIGURE 1 HERE

Starting with the comprehension stage, the teacher needs to have some knowledge of the material that she will teach to her students. In the transformation stage, the teacher must select a strategy for representing that material from her repertoire of strategies, and adapt that strategy based on their specific students’ needs. The instruction stage involves the act of teaching and possibly further adapting the strategic approach on the fly, while the assessment stage involves checking for student understanding and possible obstacles to learning (both formally and informally) in order that the teacher may further adapt their instruction to meet a students’ needs. In the reflection stage, the teacher reviews and critically analyzes the teaching episode. The progression through these stages leads to new comprehension of the subject matter material, pedagogy, and the contexts in which it is taught—in other words, PCK.

Peterson & Treagust (1995) conducted a study with pre-service primary science teachers which used Shulman’s model of pedagogical reasoning as an organizing feature. They conducted interviews and observations, and examined the journal entries and questionnaire responses of 21 pre-service teachers at the beginning and end of a problem-based learning (PBL) unit on one of two topics: magnets or air. The data were analyzed for evidence of two things: “knowledge base for teaching” (which was operationalized as science content knowledge, knowledge of learners, and curriculum knowledge), and pedagogical reasoning. The researchers specifically note that at no time during the instructional sequence (the PBL units on magnets or air) were the students given any formal instruction on the construct of knowledge base for teaching or on pedagogical reasoning. Peterson & Treagust concluded that after completing the PBL unit, the pre-service teachers were “more likely to use a pedagogical reasoning framework” and that they had “integrated their knowledge of science content, curriculum, and learners in the process” (p. 304).

Of particular note is the fact that the pre-service teachers seemed to move from more teacher-centered approaches to more student-centered approaches during the course of the PBL unit, recognizing the importance of students' prior knowledge.

The findings of this study suggest that as teachers engage in Shulman's pedagogical reasoning process, their knowledge of learners *changes* and becomes integrated with their knowledge of subject matter and representations of that subject matter. The teachers in this study adapted their teaching strategies to be more student-centered as they came to recognize the importance of their students' prior knowledge. Both of these knowledge domains (knowledge of learners and knowledge of representations of subject matter) were identified as being common to the definitions of PCK and empirical studies invoking PCK discussed above (see Table 1).

Also central to Shulman's pedagogical reasoning model of PCK development is the notion of *adaptation*. In each of the stages of the pedagogical reasoning process, the sophisticated teacher needs to consider how they will adapt their chosen teaching strategy based on the needs of their students. Hatano and Inagaki (1986), see the development process as becoming—at the most sophisticated level—an *adaptive expert*. Not only does this development involve the experience of enacting different teaching strategies and techniques, but it necessitates "an understanding of the meaning and nature" (p. 262-3) of those strategies and techniques and of their contextual relevance. For these adaptive expert teachers, "knowledge is conditionalized—it includes a specification of the contexts in which it is useful" (Bransford, Brown, & Cocking, 1999, p. 43)[3]. This idea of an "adaptational repertoire" is also supported by

---

[3] The notion of "conditionalized knowledge" came to fore empirically in our scoring moderation sessions for the first pilot test of an earlier version of the FASCI instrument. During these sessions, a consensus was reached that respondents who made conditional statements about their likely instructional strategies in various classroom scenarios represented the highest level of FA that could be identified in the data.

research on science teachers' PCK. Clermont, Borko and Krajcik (1994) found that experienced teachers have a greater adaptational repertoire than novices (see Table 1).

Shulman's model of pedagogical reasoning and Hatano and Inagaki's concept of adaptive expertise can help to define various levels of sophistication on the FASCI construct. Each of the dimensions on the FASCI are hypothesized as being continuous in nature, but are specified as having discrete levels so that FASCI responses can be scored reliably. In the next section we further describe these levels.

The FASCI Construct

The FASCI construct is composed of two dimensions: Flexible Application (FA) and Student Centered Instruction (SCI). The dimensions are closely related to the two characteristics that were identified as being common across various definitions of the construct of PCK, as discussed in the review of both theoretical and empirical work: 1) teachers' representations of subject matter and strategies for teaching the subject matter (related to the FA dimension), and 2) teachers' knowledge of students' understanding of the subject matter (related to the SCI dimension). The construct maps for each of the FASCI dimensions are shown in Figures 2 (the FA construct map) and 3 (the SCI construct map).

INSERT FIGURE 2 HERE

The FA Dimension

We consider teachers' strategic knowledge to be aspect of their pedagogical content knowledge (e.g., Abell, 2007; Magnusson, et al., 1999; Shulman, 1987). This knowledge base develops as novice teachers become more expert. With development, the novice not only gains a larger repertoire of strategies, but they also gain the ability to judge the appropriateness of various strategic approaches given the situational constraints and the ability to modify those strategies based on these constraints (e.g., Berliner, 2001; L. Bond, Smith, Baker, & Hattie, 2000; Hammerness, et al., 2005). These strategies are based in the teachers' knowledge of representations of the subject matter, which is a characteristic of pedagogical content knowledge and an essential component of the pedagogical reasoning process discussed above.

Hatano and Inagaki distinguish between routine and adaptive experts and we have appropriated this distinction in our conceptualization of the FA dimension. A teacher who is at an expert level on the FA dimension is akin to the *adaptive expert* (as defined by Hatano and Inagaki) in that she has a large "adaptational repertoire" of teaching strategies similar to expert teachers described in other studies (1994). At a slightly less sophisticated level on the FA dimension, a teacher has a repertoire of instructional strategies but chooses a strategy without consideration of relevant contextual factors (such as student understanding). The teacher can adapt a teaching strategy as needed, but does not justify her adaptation. At the lowest level on the FA dimension, a teacher has a very limited repertoire of strategies to choose from and once she has decided on a strategic approach for a given context, she does not adapt or change it. One can think of this teacher as not having progressed through the various stages of Shulman's pedagogical reasoning process, in that she has not critically analyzed her strategic approach or tailored it to the students' needs.

We define a "strategy" as a teaching act chosen purposefully to engage students with the content or skills they are expected to learn. So while all strategies are acts within the Instruction stage as presented in Shulman's model in Figure 1, not all acts necessarily constitute instructional strategies as we are defining the term. For example, the use of humorous anecdotes by a teacher can go a long way towards establishing a classroom environment that is conducive to learning, but we would not characterize this as an instructional strategy in and of itself.

Treagust (2007) distinguishes six categories of instructional strategies commonly used in science teaching:

1. Demonstrations
2. Explanations (e.g., telling students the answer, interactive explanations)
3. Questioning (e.g., dialog, debate, discussion, use of wait-time)
4. Representations (e.g., concrete, verbal, mathematical, visual, symbolic, gestural)
5. Group and Cooperative Learning (e.g., jigsaw, peer-tutoring, group investigations)
6. Inquiry (e.g., exploration, conceptual invention, conceptual expansion)

All teachers have a potential repertoire consisting of these different instructional strategies. While any given classroom activity will usually involve some combination of strategies, one will usually be most prominent. Treagust (2007) presents the categories above as existing on a continuum from "teacher-centered" to "student-centered," with demonstrations being the least student-centered and inquiry being the most student-centered. For FASCI responses, we say that an instructional strategy has been modified when the respondent articulates a shift in emphasis within a given item response from one instructional strategy to another (i.e., from a demonstration to questioning). The key criteria we use for determining whether a strategy is student-centered is whether the strategy is being used to implement a classroom activity that enables students to be actively engaged with the concepts at hand, and that provides an opportunity for student to express their developing ideas and conceptions.

INSERT FIGURE 3 HERE

The SCI Dimension

Eliciting and building on student prior knowledge is central to teaching for understanding (e.g., Bransford, et al., 1999; Fosnot, 1996; Greeno, Collins, & Resnick, 1996). By eliciting and responding to student ideas, teachers can meet the diverse needs of their students (Hammer, 1996; McDermott, 1991; Minstrell, 1991; van Zee & Minstrell, 1997). Moreover, several common ideas that students bring into the classroom and use in various contexts have been identified through research in science education (American Association for the Advancement of Science Project 2061, 1993; Driver, 1994; Driver, Guesne, & Tiberghien, 1985; Minstrell, 1991). When teachers are aware of common student ideas, they can anticipate the types of ideas that their own students might have and prepare instruction that is catered to their needs (Brown & Clement, 1989; diSessa & Minstrell, 1998; McDermott, Shaffer, & Somers, 1994; Rochelle, 1998). Whether a teacher responds to the ideas of her students in real time or over longer periods of time, when this response is used to modify instruction she is engaging in a process of formative assessment.

Formative assessment consists of goal identification, assessment, and feedback (Atkin, et al., 2001; Sadler, 1989). More specifically, assessment is formative when the information derived from the assessment informs instructional practices in order to meet identified needs of students (Black & Wiliam, 1998). The formative assessment process is *responsive* in the sense that the teacher responds to her assessment of students' knowledge states by setting intermediate goals, making instructional decisions, and providing feedback and relevant instruction (as shown in Figure 4).

In seeking to measure teachers' strategic knowledge, we are interested in the extent to which teachers' instruction involves formative assessment, or more specifically, the extent to which teachers' instruction is student-centered. Otero and Nathan (2008) demonstrated that the 61 pre-service elementary teachers in their study commonly held one of four views about the role of student's prior knowledge in their instruction. While Otero and Nathan did not hierarchically organize the four common views that they identified, three of the views they identified and described were lacking certain aspects of the formative assessment process depicted in Figure 4. With the FASCI, we hoped not only to measure differences among teachers in their views of what it means for instruction to be student centered, but also to investigate various levels of sophistication of these views. Otero and Nathan argued that the formative assessment cycle shown in Figure 4 should be applied flexibly and should be sensitive to contextual features within the classroom.

INSERT FIGURE 4 HERE

Good instruction does not *always* have to be student-centered. For example, Treagust (2007) discusses demonstrations (which are by nature very teacher-centered) as motivating for students, or used as a one part of a more student-centered approach such as a P-O-E (Predict-Observe-Explain). Further, in the formative assessment process the feedback a teacher provides might involve an activity that meets the needs of the students through a more didactic method, which could help a student move from the idea they currently have to an idea that the teacher hopes he will develop (e.g., Black & Wiliam, 1998). Hence, instruction must be flexible, in that the teacher must be able to gauge what instructional method is appropriate at a given time for a given student. This flexibility is what we sought to measure with the FA dimension.

A teacher at the expert level of SCI (level "1") recognizes the importance of knowledge of specific student learning difficulties by viewing the learning activity as a situation in which she and/or the students can find out what the students are thinking and what potential pitfalls they may be approaching. This level of thinking was identified in the definitions and uses of PCK (see Table 1) and in the pedagogical reasoning study of Peterson and Treagust—as teachers moved through Shulman's pedagogical reasoning process, they became more student-centered and aware of the value of students' prior ideas. At a novice level on the SCI dimension (level "0") the teacher views the learning activity as a non-interactive place where she represents the material to her students without any adaptation or tailoring of the material and representations to her students' needs. One can think of this situation as being very low in the pedagogical reasoning process, only beginning to engage in the transformation stage. Again, the SCI dimension is conceptualized as being continuous in nature, but previous work has revealed that at this point, only two discrete levels can be distinguished among responses and therefore scored reliably (Talbot & Briggs, 2008).

Development of the FASCI Instrument

Before setting out to develop the FASCI, we first reviewed previous attempts to measure similar constructs among science teachers. For a discussion of this review, see Talbot and Briggs (2008) and Talbot (2008). The foundation for our instrument development has been Wilson's (2005) construct modeling approach to measurement[4]. In this approach, one begins by establishing a theory for the latent construct to be measured. The instantiation of this theory is the construct map. Next, a set of items is designed to elicit information about how much or how little of the construct any given respondent possesses. A scoring rule is subsequently established for the anticipated item responses. After collecting data through pilot tests or field tests, patterns of scored responses across examinees and items are modeled statistically to make inferences about the location of a respondent (and item) on the construct map. The precision of the inference is evaluated by its associated estimate of measurement error. Each step of the construct modeling approach typically leads to iterative revisions of the construct map, items,

---

[4] This approach can also be mapped onto the "Assessment Triangle" presented in *Knowing What Students Know* (Pellegrino, Chudowsky, Glaser, & National Research Council Division of Behavioral and Social Sciences and Education Committee on the Foundations of Assessment, 2001). We start with a theory of respondent cognition (in this case, the learning progression), proceed to gather some observations of that cognition (through responses to the instrument), and use a measurement model to interpret those responses.

scoring rules, and measurement model. Indeed the results we present below obscure the many iterations and revisions that preceded them.

The construct maps presented above in Figures 2 and 3 posit two distinguishing features of teachers' strategic knowledge as they engage (or do not engage) in the formative assessment process. First, formative assessment depends on a teacher's attention to student ideas. This constitutes the student-centered instruction (SCI) dimension of a teacher's strategic knowledge (Figure 4). Second, the formative assessment process depends on a teacher's attention to student learning outcomes and use of such information to modify their strategies when these strategies do not appear to be working. This constitutes the flexible application (FA) dimension of a teacher's strategic knowledge (Figure 2). We would expect that the two dimensions in the construct map of strategic knowledge to be positively correlated since a teacher who is high on the SCI dimension is by definition one that is attentive to student conceptions and ideas, and therefore more likely to use such information to modify their instructional strategies than a teacher with a more behavioral view of teaching. As is evident in Figures 2 and 3, each dimension of the construct map is characterized by a rather small number of qualitative levels. This is not meant to suggest that more levels do not exist, only that to this point we were unable to establish additional indicators that could be used to distinguish them reliably.

The FASCI construct map represents an initial hypothesis about the factors that distinguish teachers that are more sophisticated in their strategic knowledge from those that are less sophisticated. The next step in developing an operational instrument was to create items capable of eliciting information about what teachers do (and the reasons why) when they are presented with a classroom situation.

Designing Scenario-based Items

The items on a survey instrument are the means by which information about a given respondent's location on the construct map is elicited. In the present assessment context the optimal choice of items to pose—both in terms of content and format—is not at all straightforward. Our approach has been to develop standardized items that present teachers with a variety of classroom-based scenarios. The following contextual constraint is given at the beginning of each scenario: "Please assume (unless it is otherwise specified) that you are teaching a high school course in physics, chemistry, biology, earth science or math to a class of 25-30 students." The use of these classroom scenarios is intended to make it harder for respondents to provide what they consider to be the desired answer. It is worth noting that the presentation of scenarios as a means of eliciting teacher responses is similar to the use of "key events" by Shulman's research group in their studies on teacher knowledge (described above).

For each scenario, we provide three prompts that allow teachers to elaborate on their instructional decisions. The general structure for each item is shown below.

Step 1: Present a classroom scenario that involves a specified learning activity. The scenarios should vary to the extent that they appear more student-centered or teacher-centered.

Item Prompt (a) asks: How might this activity facilitate student learning?

Step 2: Present information hypothetically elicited from the learning activity posed in the scenario that represents a potential obstacle to student learning.

Item Prompt (b) asks: Describe both what would you do and what you would expect to happen as a result.

Item Prompt (c) asks: If the approach you described above in (b) didn't produce the result(s) you anticipated by the end of that class session, what would you do in the next class session?

The item below illustrates how the template above is applied:

Students are working in groups of four to discuss a conceptual question you provided them at the beginning of class.

a) How might this activity facilitate student learning?

As the activity proceeds, one group gets frustrated and approaches you—they've come up with two solutions but can't agree on which one is correct. You see that one solution is right, while the other is not.

b) Describe both what would you do and what you would expect to happen as a result.

c) If the approach you described above in (b) didn't produce the result(s) you anticipated by the end of that class session, what would you do in the next class session?

While we have developed a total of 15 different scenarios, the pilot testing of the FASCI that we discuss below contains only the following five:

Scenario 1) Students are working in groups of four to discuss a conceptual question you provided them at the beginning of class.

Scenario 2) You are working out an example problem up on the board.

Scenario 3) You have just finished giving a lecture on a complicated topic.

Scenario 4) You have given your students a quiz to assess their understanding of a difficult topic.

Scenario 5) In talking with one of your students you discover that they have a misconception[5] about a central topic presented in that week's class. You attempt to address the misconception by having a one-on-one conversation with the student.

Many of the scenarios created for the FASCI instrument were inspired by the actual experiences recounted to us in interviews with LAs. All scenarios were written by a team of

---

[5] Although we recognize the potentially problematic nature of using the term "misconception" rather than "alternative conception" or "prior ideas," we chose to use it in this particular item because it reaches a larger audience of respondents when interpreted in its colloquial meaning.

faculty, graduate students, and K-12 teachers with expertise in both measurement and science education. In addition to these scenario-based items, a set of background questions are included about a respondent's academic background and previous teaching experiences[6].

Scoring Rules for FASCI Items

We developed two separate rubrics with rules for scoring every set of three item responses to each scenario on the FA and SCI dimensions. These scoring rules are based on the construct maps shown in Figures 2 and 3. SCI scores are based solely upon responses to prompt (a), while FA scores are based on the combination of responses to prompts (b) and (c). A response is given an SCI score of "1" rather than "0" only if there is some explicit evidence that the teacher views the potential for learning in this activity from a student-centered standpoint. In such responses, the classroom activity is viewed as an opportunity for the teacher to interact with the students, for the students to interact with one another, or for individual students to express or defend their ideas.

After prompt (a), new information is provided in prompt (b) about a potential obstacle to learning under the initial activity. After describing this obstacle, prompt (b) asks the respondent what he or she would do and what he or she would expect to happen as a result. This marks the first opportunity for a teacher to choose an instructional strategy beyond the one provided in the initial scenario. We do not however, consider this response a modified strategy, but rather a baseline choice of strategy. It is in prompt (c)—when the respondent is told that the approach he or she described in (b) has not produced the expected result—that we look for evidence that a teacher can modify his or her baseline instructional strategy. If the response indicates a modified instructional strategy in prompt (c), the FA score moves from a "0" to "1." If the respondent then

---

[6] We have found that it takes the average respondent about 30 minutes to complete the most recent version of the FASCI.

qualifies the modification with one or more conditional statements (typically based on contextual features), then the FA score moves from a "1" to a "2." At this highest score level we are trying to capture the type of teacher that recognizes that there will always be a context when a given instructional strategy, no matter how carefully considered, may fail. Such teachers, we suspect, will tend to answer prompt (c) with "it depends" and then explain why.

Each scenario on the FASCI instrument results in a separate item score for each dimension. The pilot test version of the FASCI contained 5 scenarios; hence there were 5 items with a maximum raw score of 5 points for the SCI dimension, and 5 items with a maximum raw score of 10 points for the FA dimension. These raw scores were subsequently modeled using item response theory which we describe below.

Modeling of FASCI Item Responses

Once FASCI item responses have been scored, our aim is to relate inferences about the score distributions within and across respondents back to the underlying construct map. In doing so we need to be able to characterize which respondents are "lower" or "higher" on the FASCI construct map, and which items are easier or harder to answer in a manner we would score as sophisticated. We also need to characterize the precision with which we can draw inferences about the score differences among individual purposes. To accomplish these purposes we model each dimension of the FASCI separately using Item Response Theory (IRT; Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991; Lord, 1980). This allows us to place both persons and items on the same scale[7]. The location of one can then be interpreted probabilistically relative to the location of the other using a graphical format known as a Wright Map (Wilson, 2005, p. 90). A Wright Map is the empirical analog to the qualitative distinctions

---

[7] For a more detailed description of this statistical modeling, see Talbot and Briggs (2008) in which we describe in detail the IRT models used with the FASCI data.

that were hypothesized as part of our FASCI construct map and scoring rules. For our purposes in this paper, the key outcome of interest after modeling FASCI scores is the extent to which different groups of respondents can be reliably distinguished on the FA and SCI constructs.

Evidence for Distinguishing FASCI Respondents

Sample

In our most extensive pilot test of the FASCI instrument to date (administered during the Fall semester 2007), 65 individuals responded to the 5 scenario-based tasks. A description of the 65 respondents is shown in Table 2. It is this sample from which all of the quantitative evidence and much of the qualitative evidence below is presented.

INSERT TABLE 2 HERE

We purposefully chose our sample in order to gather a heterogeneous set of respondents. We hypothesized that very novice teachers (such as LAs) would have a limited repertoire of instructional strategies (and therefore lower FA scores) than more experienced teachers such as practicing K-12 teachers or Faculty Experts. Similarly, we expected novices such as graduate students in the STEM disciplines to be less student-centered in their instructional approaches (and therefore have lower SCI scores) as compared to more experienced K-12 teachers and Faculty Experts.

Quantitative Evidence

Responses to the FASCI items were independently rated by three trained raters using detailed scoring guides and supporting documents that had evolved from earlier pilot tests and their associated scoring moderation sessions. The percent agreement between raters on FA scores across five tasks ranged from 61% to 91% with an average of about 79%. The percent agreement on SCI scores across five tasks ranged from 45% to 95% with an average of about 67%. The variability in the score agreement across tasks suggests that considerable work will need to be done to improve the training of raters, although this problem is mitigated to the extent that all raters score the same respondents and all discrepant scores are resolved. This was the approach taken for the scoring reported here. The frequency distribution of final item scores by dimension for all 65 respondents is provided in Table 3. Note that there is missing data for some dimensions on some items. Approximately 5% of the surveys were not entirely completed.

INSERT TABLE 3 HERE

We modeled the two dimensions of the FASCI separately using the Partial Credit Model, and estimated associated item and person parameters using the software ConQuest (Wu, Adams, & Wilson, 1997). In general, the match between the responses predicted by the model and the item scores observed was quite good[8]. We examined the Wright Maps for each dimension, which display respondents' location related to item difficulty. The score scale on the Wright Map is represented in logit values that range from negative to positive; these values mark the location of both respondents and items on a common scale. The mean person sophistication estimate for the FA dimension is just below -1 logit, which corresponds to a raw score of 3.8 points (out of 10 possible); for the SCI dimension, mean person ability is about 0 logits, which corresponds to a raw score of 2.7 points (out of 5 possible). For a more detailed description and display of the Wright Maps, see Talbot and Briggs (2008).

One noteworthy finding that the FA Wright Map helps make apparent is the large gap between the difficulty thresholds for each item. This is indicative of the small number of respondents (see Table 3) who give a response to prompt c that takes context into account when discussing their modification of an instructional strategy. Hence for most respondents, it was

---

[8] The fit of the category characteristic curves for each item was evaluated using Weighted Mean Square Error fit statistics (T. G. Bond & Fox, 2001; Wilson, 2005; Wright & Masters, 1982)

relatively easy to give a response scored in the middle FA category (a "1") relative to the lowest (a "0"), but it was considerably harder to give a response that was scored in the highest FA category (a "2"). An examination of the SCI Wright Map reveals that two out of five items (items 1 and 5) were relatively easy for respondents. The easiest item in which respondents demonstrate student-centered thinking is item 1, in which all but three respondents received a score of 1. This is not entirely surprising, as we would expect the scenario of "students working in groups" to be an easy situation for an individual to conceive of as student-centered or as an opportunity for interactive teaching. The next easiest was item 5, in which the scenario states that "you are discussing a misconception with a student in a one-on-one conversation." Again, we would expect most respondents to conceive of this situation as an opportunity for interactive teaching. The most difficult SCI item is the scenario in which the respondent is told that they have "just finished giving a lecture on a complicated topic" (item 3). Again, we would expect this scenario which presents a formal, teacher-led presentation would be a more difficult one to conceptualize as a student-centered instructional activity. Overall, the SCI items span a range of difficulty which includes the ability estimates of all respondents in the pilot sample.

The locations of respondents on the FA and SCI continua are estimates, and as such they contain some degree of measurement error. This measurement error can be quantified for each respondent by what is known as the standard error of measurement (SEM). If scores from the FASCI are to be used for summative purposes, then it is important that the SEM is relatively small. Across the range of estimated respondent locations on both the FA and SCI dimensions, the SEM is about 1 logit[9]. To evaluate the magnitude of the SEM, we can contrast it to the range of observed respondent locations. For the FA dimension this range is about 7 logits (from -4

---

[9] In IRT models, the SEM actually varies as a function of ability, so the SEM of 1 reflects the typical value found. The range of observed SEM values is between 1.05 and .84 for the FA dimension, and between 1.2 and .90 for the SCI dimension.

logits to 3 logits); for the SCI dimension this range is much less, about 4.2 logits (from -2 logits to 2.2 logits). This implies that classification of respondents into statistically distinguishable categories is currently more plausible on the basis of FA scores than it is for SCI scores. Using a more formal approach described by Wright & Masters (1982, p. 105-106) one could conclude that there are at least 3 statistically distinct groupings (or person "strata") of respondents on the basis of estimated locations on the FA dimension, but only 1 on the SCI dimension. A traditional way to summarize the proportion of score variability that is "real" and not due to measurement error is with a reliability coefficient. The reliability coefficients for the FA and SCI dimensions are 0.71 and 0.47 respectively. In a more recent pilot test (Fall semester 2008), initial analyses show reliability coefficients for the FA and SCI dimensions of 0.71 and 0.73 respectively. Note that in this more recent pilot, the number of scenarios given to respondents was six rather than five, and only two scenarios were common between the two pilot tests.

To what extent are the distributions of respondent sophistication estimated on the FA and SCI continua consistent with the discrete levels that were hypothesized in the FA and SCI dimensions of the FASCI construct map? Based on the Fall 2007 pilot test, the answer for the SCI dimension is relatively straightforward at this point: FASCI items are not yet capable of reliably distinguishing respondents on this dimension—almost half of the observed variability on SCI items can be attributed to measurement error. Because of this, there is little point in trying to map SCI scores back to the SCI dimension of the FACSI construct map. In the discussion section, we consider ways that SCI reliability might be increased, taking into account the preliminary results of our more recent pilot test. On the other hand, at least from a statistical perspective, it is possible to reliably distinguish three levels on the FA continuum. However, the determination of whether these are the same levels described in Figure 2 requires subjective judgment. This judgment might be analogous to the standard setting activities conducted to establish cut-points for the scores on state-level standardized tests considered to represent, for example, "proficient" performance. If such an approach were to be taken, one would need to pick two cut-points on the FA Wright Map that distinguish respondents whose collective pattern of item responses was reflective of levels 0, 1 or 2 on the FA dimension. For the sake of illustration, the FA cut-point differentiating a 0 from 1 might be set at -2 logits, and a 1 from a 2 might be set at 2 logits. Another possibility, if the underlying sample of teachers had been sampled such that they were representative of some well-defined population, would be to establish the cut-points on the basis of norm-referenced criteria.

There are two hypotheses about FASCI score interpretations that we can begin to evaluate with our pilot test data. First, we have assumed that the scores for each dimension would be positively correlated. In fact, the correlation between the dimensions is 0.59, a value that actually understates the magnitude of the relationship because it is attenuated by measurement error.

Second, we have assumed that higher scores on the FASCI dimensions are indicative of substantive distinctions among teachers. Recall that there were 5 subgroups to consider:

1. Learning Assistants (*n* = 43)
2. Noyce Fellow (*n* = 2)
3. Practicing K-12 Teachers (*n* = 8)
4. Faculty Experts (*n* = 3)
5. Graduate Students (*n* = 9)

While these results are hardly conclusive given the small and limited sample we have gathered, they are suggestive nonetheless. Because of the observed unreliability of SCI scores, we focus on group comparisons for the FA dimensions. We note that the relatively inexperienced LAs (with a mean teaching experience of 1.2 years) score lower on the FA dimension than the more experienced Noyce Fellows (mean teaching experience of 5 years) or the Faculty Experts (mean teaching experience of 30 years). Because Noyce Fellows take on increasing levels of responsibility, and because they are enrolled in education courses in which they have many hours per week working with K-12 students, we would expect them to score higher on the FA scale than the LAs. One surprising finding is that our sample of practicing K-12 teachers (mean teaching experience of 15.8 years) also scored rather low on the FA dimension. In fact, they are located at essentially the same place as the Graduate Student teaching assistants, who we expected to be low due both to their lack of experience (mean of 1.7 years) and their lack of teacher education. We had hypothesized that the FA scores of K-12 teachers would be higher based in part on their extensive classroom experience. One possible explanation for their low scores could have to do with their teacher education and/or induction programs; another possible explanation is that our interpretation of these scores in invalid. The best way to evaluate this would be to observe the same teachers in their classrooms, a validation activity we are currently undertaking. Preliminary results from these observations are discussed below.

Observations of Practice

In order to guide our observations of practice, we developed the FASCI observation protocol (available from first author upon request). This protocol is directly linked to the FASCI construct, and prompts the observer to check if teachers are performing an action once or more than once during the observed lesson. Specifically, the protocol asks "Does the teacher elicit students' ideas or current conceptions about the topic?" (aligned with the SCI dimension), and whether or not they employ specific teaching strategies from a list based on the work of Treagust (2007) (aligned with the FA dimension). Perhaps most importantly with respect to the FA dimension, observers note if anything happened during the class that appeared to be unexpected by the teacher and if so, what the teacher did in response. This is where we capture information about the teacher's adaptive expertise. In the post-observation interview, the teacher is also directly asked if "anything happened in class today that you did not expect to happen" and further probed if they answer in the affirmative.

The FASCI observation protocol is currently being used with teachers who are part of a longitudinal research project studying the effects of the LA program (Talbot, et al., 2009). Teachers in their first three years of teaching are given the FASCI once per year, and observed teaching three times per year. The FASCI observation protocol (along with the Reformed Teaching Observation Protocol (RTOP; Sawada, et al., 2002)) is being used in each of those observations. Preliminary results with respect to the FA dimension show that teachers use many different teaching strategies, but that evidence of being *adaptive* (i.e., responding to an unexpected event) is rare. This is consistent with most of their FASCI scores on the FA dimension, which describe them as having a large repertoire of teaching strategies which is not often applied in a conditionalized way. With respect to the SCI dimension, most of the teachers observed seem to actively elicit students' current ideas, which is consistent with their FASCI scores. However, as mentioned previously the SCI scores resulting from the first FASCI pilot test were very reliable. In the latest pilot, SCI score reliability increased markedly. One avenue for us to explore is the relationship between these more reliable SCI scores and data from future applications of the FASCI observation protocol.

Cognitive Interview Data

A subset of respondents (*n* = 9) volunteered to participate in cognitive interviews after having taken the FASCI. These interviews are ongoing, and are coupled with another FASCI validation study that is currently being conducted (the development of the physics-specific FASCI, discussed below). Our purpose in conducting these structured interviews is to find out what aspects of the survey the respondents find difficult, easy, and frustrating. Further, we explicitly ask them to talk us through their thought processes as they are formulating their responses to the scenario-based items. We designed the latter aspect of each interview to not only give us insight into thoughts processes, but also to help us define in a more fine-grained way what the respondents actually think they would do in the situation.

The most notable result from the cognitive interviews is that in general, respondents seem to be formulating or invoking their own context in which to respond to the item prompts. This lack of context was seen as a difficulty by some respondents:

> "I think the difficult part of some of these [scenario-based items] was having to speak so generally about them, since I didn't know what this group of four specifically was working on. I felt like a lot of my answers were kind of maybe equally vague, I suppose." (ID: 3475220)

Whether or not this context-neutrality poses a serious threat to the validity of FASCI score interpretations is the subject of a validity study that is currently being undertaken. We are conducting an experiment in which we are administering either the current version of the FASCI or a parallel, physics-specific version at random to a population of pre-service science teachers. This study is further explained in the discussion section.

When asked questions about their responses to prompts that were scored on the FA dimension, respondents often discussed a specific strategic approach in response to the potential obstacle presented, and did not directly cite contextual conditions which would bear on their choice of approach. This is consistent with their written responses—most seem to have a developing repertoire that they invoke without articulated justification—and with preliminary results from the observation protocols.

With respect to the SCI dimension, one respondent talked about active student engagement more than his written responses would indicate. For example, on the item about "having given the students quiz on a difficult topic" (scenario 4), this respondent's written response would be scored as SCI = 0. However, in the cognitive interview, he elaborated on the active role the student must take in interpreting their quiz performance and in taking action to better understand the topic. This discord between written and discussed response needs to be explored more as we continue to build a validity argument for FASCI score interpretations.

Discussion

The FASCI instrument was designed to measure teachers' strategic knowledge, which we have operationalized with the FA and SCI dimensions as the way teachers conceptualize the application and revision of instructional strategies in classroom learning environments. The instrument was designed for two different possible uses, one formative and one summative. In both cases the objective was to provide users of the instrument with two scores that could be used to make meaningful distinctions about the "sophistication" of a given teacher's strategic knowledge. To accomplish this we developed (through iterative revision) a construct map, and then created scenario-based items that could be scored on the basis of this construct map. Finally, we pilot-tested the FASCI using a sample of 65 teachers with varying levels of experience, and used Wright Maps of each dimension to characterize the empirical results. Our findings to this point indicate that the FASCI does not yet measure respondents reliably enough to support its use in high-stakes contexts. There is, however, some evidence to suggest that meaningful distinctions can be made on the FA and SCI dimensions, and that observed group orderings for LAs, Noyce Fellows, Graduate Students and Faculty Experts are as we had expected. Also, evidence from observations of practice and cognitive interviews support these distinctions. Furthermore, we have some evidence, on the basis of focus group interviews and open-ended comments at the end of the survey, that the act of taking the FASCI and discussing responses can serve as rich instructional activity in its own right. For such low-stakes uses, issues of measurement error are much less of a concern. Clearly, for high-stakes uses much more work needs to be done to both improve the reliability of FASCI scores. In addition, whether it is to be used for formative or summative purposes, more research must be done to validate the FASCI score interpretations implied by the underlying construct map. Below we briefly describe these future research directions.

To increase score reliability, principally on the SCI dimension, we are focusing our efforts on the development of a partially-constrained version of the FASCI. We are currently piloting this new version of the FASCI in which the responses to each of the three prompts within a single scenario are multiple choice. There are two unique aspects to this design: 1) respondents are "branched" to an open-ended question if they choose the multiple choice category "it depends" in response to the third prompt ("If the approach you chose above didn't produce the results…"), and 2) respondents are asked to rate their degree of confidence in answering the multiple choice items, and are asked to tell us (in an open-ended fashion) what would make them more confident in their choices. We expect the constraint of answer choices to boost reliability of the SCI dimension, partly because the answer choices are based on three levels of sophistication (based on the work of Otero & Nathan, 2008) rather than our current two level scoring. Further, by leaving some open-ended responses we will still be able to distinguish the most sophisticated respondents as they have the opportunity to further justify their choices and confidence.

We are also in the process of conducting three validity studies. In the first we are continuing to conduct and analyze cognitive interviews with respondents after they have completed the FASCI. We hope to further use evidence from these interviews to gain insights about the cognitive processes invoked when teachers respond to FASCI prompts, and the extent to which they are consistent with our scoring criteria. In the second study we are continuing to observe a small subset of FASCI respondents into their classrooms in order to evaluate whether the flexibility and student-centeredness implied by their FASCI scores is consistent with what we observe. In the third study we challenge an underlying assumption of the FASCI that the two-dimensional construct is in some sense content neutral. At present, each of the FASCI scenario items is in some sense content free—that is each does not depend on the actual content being taught. This was done to make the FASCI scenarios general enough such that teachers in different science content areas could be given the same items. Of course, one might intuitively expect that a teachers' strategic knowledge does, in fact, depend upon the specific content they are teaching. In this validation study, we are administering the current (content-neutral) version of the FASCI or a newly developed, physics-specific version of the FASCI at random to a population of pre-service teachers. In this new version, each scenario is couched within a specific physics topic in order to constrain the context within which the respondent is framing their response.

We now turn to issue of whether science teacher's strategic knowledge be conceptualized as a learning progression. The construct map underlying the FASCI is quite consistent with the general definition of a learning progression. One could easily define successively more sophisticated ways of thinking about teaching that follow one another as teachers gain expertise through the combination of FA and SCI levels provided in Figures 2 and 3. However, defining a learning progression is one thing; locating respondents on this progression with quantitative measures and validating these measures is another thing altogether. The construct map we developed for the FASCI was directly influenced by the constraint that levels of sophistication can be distinguished by factors that could be elicited and observed through a survey instrument. If the construct map were purely of theoretical interest, then the SCI dimension would have been more likely to resemble something like a hierarchically ordered version of teachers' conceptions of formative assessment empirically determined by Otero and Nathan (2008). This raises what we view as an important issue for those conceptualizing learning progressions but making no concurrent attempt to measure what they are conceptualizing: if it cannot be measured, does it really exist? Further, we believe that this concurrent measurement should be part of a continuous, iterative approach (much like instrument development itself) which continually informs the learning progression.

While the construct map underlying the FASCI *could* theoretically be conceptualized as a learning progression, at present we believe it falls short. There are two principal reasons for this. First, we are uncertain about the time frame over which teachers "learn" strategic knowledge. Does this happen over a short period of time such as during a teacher education program? Does it continue through an induction program and throughout a teacher's career? Is strategic knowledge even sensitive to instruction? In our view an implicit assumption of any learning progression is that given sufficient instruction and time, it should be possible to move any student from the lowest to the highest level of the progression. This would be a difficult assumption to validate for a learning progression of strategic knowledge; at a minimum, we would need to gather longitudinal data to support the presumption such changes in strategic knowledge are occurring for teacher populations such as the LAs participating in the STEM-TP program. Of course, this presents a chicken and egg problem—if no change were to be found on FASCI scores over some defined period of time, should we infer that strategic knowledge is not sensitive to instruction, or that the FASCI instrument is not capable of measuring such change? A second problem with viewing the FASCI construct map as a learning progression is that at present we have made no connection between a respondent's presumed location on the construct and the instructional activities that would be recommended to move them to a higher location. That is, imagine that we could establish that a teacher conceptualizes most instructional strategies in a student-centered way (i.e., SCI = 1), and is flexible in how she adapts these strategies, but often fails to take contextual factors into account in the process (i.e., FA = 1). If this teacher was an LA in the STEM-TP program, what should be done with this information? How will it be used to further increase her level of strategic knowledge? We cannot currently answer these questions, and until we can do so, it may be a stretch to interpret FASCI scores as manifestations of a learning progression.

However, that said we do contend that taking a measurement-oriented approach to defining a learning progression is warranted. In this approach, we first develop a theoretically and empirically-based measurement tool that elicits information which can be scored reliably. Coupled with qualitative data from (in this case) observations of practice and interviews, the quantitative data can help to illuminate the underlying learning progression and further inform construct development. In this way we are not presuming that all teachers progress on the path defined by our construct, only that we can reliably place them on that construct and not on the much more detailed learning progression.

We conclude this paper by emphasizing the point that establishing something akin to a construct map is a necessary, but not sufficient condition to call something a learning progression. With enough time and energy, people with expertise in a given content domain can establish a reasonable hypothesis for "successively more sophisticated ways of thinking about an idea that follow one another as students learn." But in our view a learning progression implies more than this. A next step is to establish an empirical basis that validates the hypothesized progression, and doing so will almost always require the development of associated assessment instrumentation. The systematic approach described here to develop the FASCI instrument provides an example. The final step in establishing a learning progression is to establish them as tools for formative assessment. Where are teachers located on the progression, and what instructional course of action does this suggest? When these three pieces of development are put together, a learning progression becomes the combination of theory and action, with the action providing the empirical basis for validation and/or revision of the initial theory.

References

Abell, S. K. (2007). Research on Science Teacher Knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 1105-1149). Mahwah, New Jersey: Erlbaum.

American Association for the Advancement of Science Project 2061 (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Atkin, J. M., Black, P., & Coffey, J. (2001). *Classroom Assessment and the National Science Education Standards*.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*, 463-482.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Bond, L., Smith, T., Baker, W. K., & Hattie, J. A. (2000). *A distinction that matters - why national teacher certification makes a difference*. Arlington, VA: National Board for Professional Teaching Standards.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model : fundamental measurement in the human sciences*. Mahwah, N.J.: L. Erlbaum.

Bransford, J., Brown, A. L., & Cocking, R. R. (1999). *How people learn: brain, mind, experience, and school*. Washington, D.C.: National Academy Press.

Brown, J., & Clement, J. (1989). Overcoming Misconceptions via Analogical Reasoning: Abstract Transfer versus Explanatory Model Construction. *Instructional Science, 18,* 237-261.

Clermont, C. P., Borko, H., & Krajcik, J. S. (1994). Comparative study of the pedagogical content knowledge of experienced and novice chemical demonstrators. *Journal of Research in Science Teaching, 31*(4), 419-441.

Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical Content Knowing: An Integrative Model for Teacher Preparation. *Journal of Teacher Education, 44*(4), 263-272.

diSessa, A. A., & Minstrell, J. (1998). Cultivating Conceptual Change with Benchmark Lessons. In J. G. Greeno & S. Goldman (Eds.), *Thinking Practices in Mathematics and Science Learning* (pp. 155-187). Mahwah, NJ: Lawrence Erlbaum Associates.

Driver, R. (1994). *Making sense of secondary science : research into children's ideas*. London ; New York: Routledge.

Driver, R., Guesne, E., & Tiberghien, A. (1985). *Children's ideas in science*. Milton Keynes [Buckinghamshire] ; Philadelphia: Open University Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates.

Fosnot, C. T. (1996). *Constructivism : theory, perspectives, and practice*. New York: Teachers College Press.

Greeno, J. G., Collins, A., & Resnick, L. B. (1996). Cognition and Learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 15-46). New York: Macmillan.

Grossman, P. (1990). *The Making of a Teacher: Teacher Knowledge & Teacher Education*. New York: Teachers College Press.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Concepts, Models, and Features *Fundamentals of Item Response Theory* (pp. 7-31). Newbury Park, CA: SAGE.

Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role in education research. *American Journal of Physics, 64*(10), 1316-1325.

Hammerness, K., Darling-Hammond, L., Bransford, J., Berliner, D. C., Cochran-Smith, M., McDonald, M., et al. (2005). How Teachers Learn and Develop. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing Teachers for a Changing World: What teachers should learn and be able to do* (pp. 358-389). San Francisco: Jossey-Bass.

Hashweh, M. Z. (1987). Effects of subject-matter knowledge in the teaching of biology and physics. *Teaching and Teacher Education, 3*(2), 109.

Hatano, G., & Inagaki, K. (1986). Two Courses of Expertise. In H. Stevenson, H. Azuma & K. Hakuta (Eds.), *Child Development and Education in Japan* (pp. 262-272). New York: Freeman.

Lee, E., Brown, M. N., Luft, J. A., & Roehrig, G. H. (2007). Assessing Beginning Secondary Science Teachers' PCK: Pilot Year Results. *School Science and Mathematics, 107*(2), 52-61.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N.J.: L. Erlbaum Associates.

Magnusson, S. J., Krajcik, J., & Borko, H. (1999). Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education* (pp. 95-132). Dordrecht, The Netherlands: Kluwer.

McDermott, L. C. (1991). What we teach and what is learned: Closing the gap. *American Journal of Physics, 59*(4), 301-315.

McDermott, L. C., & Shaffer, P. S. (2002). *Tutorials in Introductory Physics*. Upper Saddle River, NJ: Prentice Hall.

McDermott, L. C., Shaffer, P. S., & Somers, M. (1994). Research as a Guide for Curriculum Development: An Illustration in the Context of Atwood's Machine. *American Journal of Physics, 62*(1), 46-55.

Minstrell, J. (1991). *Facets of Students' Knowledge and Relevant Instruction.* Paper presented at the Research in Physics Learning Workshop at the University of Bremen: Theoretical Issues and Empirical Studies, Kiel, Germany: Institute fur die Padagogik der Naturwissenshaften.

National Assessment Governing Board, WestEd, & Council of Chief State School Officers (2007a). *Science Assessment and Item Specifications for the 2009 National Assessment of Educational Progress (pre-publication version).*

National Assessment Governing Board, WestEd, & Council of Chief State School Officers (2007b). *Science Framework for the 2009 National Assessment of Educational Progress (pre-publication version).*

National Research Council Committee on Science Learning Kindergarten Through Eighth Grade., Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school : learning and teaching science in grades K-8*. Washington, D.C.: National Academies Press.

National Science Teachers Association (2003). *Standards for Science Teacher Preparation*.

Otero, V. (2006). *Recruiting talented mathematics and science majors to careers in teaching: A collaborative effort for k-16 reform.* Paper presented at the National Association for Research in Science Teaching, San Francisco.

Otero, V., & Nathan, M. J. (2008). Preservice Elementary Teachers' Views of Their Students' Prior Knowledge of Science. *Journal of Research in Science Teaching, 45*(4), 497-523.

Pellegrino, J. W., Chudowsky, N., Glaser, R., & National Research Council Division of Behavioral and Social Sciences and Education Committee on the Foundations of Assessment (2001). *Knowing what students know : the science and design of educational assessment*. Washington, DC: National Academy Press.

Peterson, R., & Treagust, D. (1995). Developing Preservice Teachers' Pedagogical Reasoning Ability. *Research in Science Education, 25*(3), 291-305.

Putnam, R. T., & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have To Say about Research on Teacher Learning? *Educational Researcher, 29*(1), 4-15.

Rochelle, J. (1998). Beyond Romantic versus Skeptic: A Microanalysis of Conceptual Change in Kinematics. *International Journal of Science Education, 20*(9), 1025-1042.

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science, 18*(2), 119-144.

Sanders, L. R., Borko, H., & Lockard, J. D. (1993). Secondary science teachers' knowledge base when teaching science courses in and out of their area of certification. *Journal of Research in Science Teaching, 30*(7), 723-736.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., et al. (2002). Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics, 102*(6), 245-253.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review, 57*(1), 1-22.

Talbot, R. M. (2008). *Measuring Science Teacher Knowledge: Domain-General or Domain-Specific?* (Research proposal). Boulder, CO: University of Colorado at Boulder.

Talbot, R. M., & Briggs, D. (2008, March 27). *Measuring the Pedagogical Sophistication of Science Teachers using the Flexible Application of Student-Centered Instruction (FASCI) Instrument.* Paper presented at the American Educational Research Association Annual Meeting, New York, NY.

Talbot, R. M., Otero, V. K., Finkelstein, N. D., Gray, K. G., Webb, D. C., Moin, L. J., et al. (2009, April 19). *A Longitudinal Study on Pedagogical Content Knowledge: Synthesizing Research on Content, Pedagogy, and Practice.* Paper presented at the Annual Conference of the National Association of Research of Research in Science Teaching, Garden Grove, CA.

Tamir, P. (1988). Subject Matter and Related Pedagogical Knowledge in Teacher Education. *Teaching and Teacher Education*, *4*(2), 99-110.

Treagust, D. (2007). General Instructional Methods and Strategies. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 373-391). Mahwah, NJ: Lawrence Erlbaum Associates.

van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing Science Teachers' Pedagogical Content Knowledge. *Journal of Research in Science Teaching, 35*(6), 673-695.

van Zee, E., & Minstrell, J. (1997). Using Questioning To Guide Student Thinking. *Journal of the Learning Sciences*, *6*(2), 227-269.

Wilson, M. (2005). *Constructing measures : an item response modeling approach*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Wilson, M., & Bertenthal, M. W. (2006). *Systems for state science assessment*. Washington, DC: National Academies Press.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis. Rasch Measurement*.

Wu, M. L., Adams, R. J., & Wilson, M. (1997). ConQuest Generalised Item Response Modeling Software: Australian Council for Educational Research.

Table 1.

Summary of Empirical Studies on Science Teacher PCK

Study	Summary of Methods	Findings	Aspects of PCK central to research
Hashweh (1987)	Studied how biology (*n* = 3) and physics (*n* = 3) teachers planned for teaching in and out of their area of science expertise.	Within their own fields, teachers were more likely to know what was problematic for students and how to deal with those concepts instructionally.	Transformation of science subject matter for teaching. Responding to student learning difficulties.
Sanders, Borko, and Lockard (1993)	Studied science teachers (*n* = 3) planning, teaching, and reflecting when in and out of their content areas.	Participants were "more expert" when teaching in their area of expertise, in that they were "very selective in their use of instructional and management" and were able to help students "avoid pitfalls." However, Experienced teachers could draw on their pedagogical knowledge when they lacked the content knowledge. Certain strategies for science teaching.	Knowledge of strategies for teaching both specific and general science content. Knowledge of common student areas of difficulty.
Clermont, Borko, and Krajcik (1994)	Studied novice (*n* = 7) and experienced (*n* = 5) "chemical demonstrators" PCK with respect to using demonstration strategies.	Experienced teachers have a greater "adaptational repertoire" and potential student learning difficulties.	Knowledge of specific teaching strategies for specific science topics Growth/change in participants' PCK (using Shulman's model of pedagogical reasoning)
van Driel, Verloop and de Vos (1998)	Teachers (*n* = 12, all had at least 5 years experience) participated in workshop on chemical equilibrium. Workshop was audio recorded.	Specific elements of PCK for chemical equilibrium exist. PCK is a unique domain of knowledge "to be discerned from general pedagogical knowledge and from subject matter knowledge."	Knowledge of representations of subject matter. Understanding of specific learning difficulties and student conceptions.
Lee, Brown, Luft, and Roehrig (2007)	Developed a rubric to characterize the PCK of beginning science teachers (*n* = 24) from different induction programs.	Beginning science teachers do not "have" much PCK. PCK develops over time. The "elements of PCK are interwoven."	Knowledge of instructional strategies. Knowledge of student learning.

Comprehension	**Transformation**	**Instruction**	**Assessment & Evaluation**	**Reflection**	**NEW Comprehension**
Teacher's content knowledge of material to be taught	1. Preparation 2. Representation 3. Instructional Selections (from an array of teaching methods and models) 4. Adaptation (taking into account differences in learning styles) 5. Tailoring of Adaptation (needs of specific students) "Fitting the represented materials to the experiences of the students"	Variety of teaching "acts" • Management • Presentations • Interactions • Group Work • Discipline • Humor • Questioning • Inquiry	• Checking for understanding • Looking for misconceptions or other obstacles to learning • Evaluating one's own performance and adjusting for different circumstances.	• Reviewing • Reconstructing • Re-enacting • Critically analyzing one's teaching	Deeper understanding of both content and pedagogy

Figure 1. Shulman's Model of Pedagogical Reasoning

Level		Respondent Characteristics
Increasing FA	2	• The teacher has repertoire of strategies that can be used to facilitate student learning within a given class session. • If the teaching strategy comprised of these acts is not producing the desired result, sometimes it can be modified. • The teacher recognizes that the choice of a class activity and associated teaching strategy will depend upon variables specific to the classroom context.
	1	• The teacher has a repertoire of strategies that can be used to facilitate student learning within a given class session. • If an activity based on a particular teaching strategy is not producing the desired result, the activity can be modified by selecting a different strategy.
	0	• The teacher has a limited repertoire of strategies. • Once a particular activity has been selected for a class session, it is not easily modified with a different strategy.

Figure 2. Construct map for the Flexible Application (FA) dimension

Level		Respondent Characteristics
Increasing SCI	1	• Discussion of interactive teaching which would be *observable* to the teacher or to an outside "other." Teacher $\leftrightarrows$ Students and/or Students $\leftrightarrows$ Students
	0	• No discussion of interactive teaching • Teacher primarily views classroom activity as way to help students make sense of new ideas. Information goes from teacher to student. Teacher → Students

Figure 3. Construct map for the Student-Centered Instruction (SCI) dimension

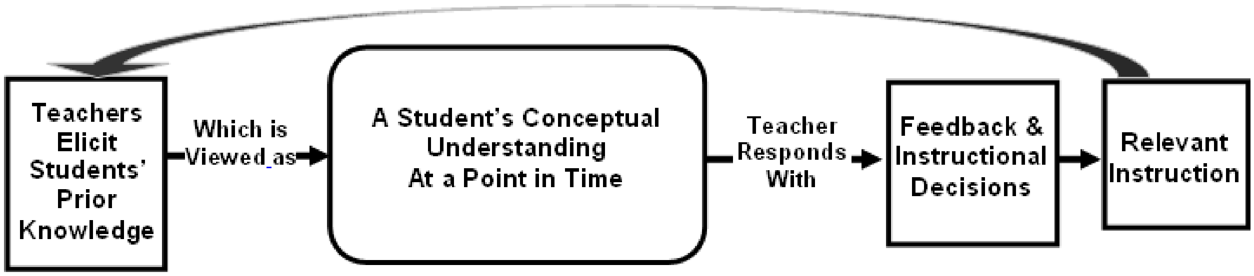

Figure 4. A cyclic model of formative assessment (from Otero & Nathan, 2008)

Table 2. Descriptive Characteristics of the FASCI Pilot Test Sample

		LA (*n* = 43)	Noyce (*n* = 2)	K-12 (*n* = 8)	Faculty (*n* = 3)	Grad Students (*n* = 9)	Entire Sample (*N* = 65)
Mean Age (SD)		20.33 (2.50)	22.5 (0.71)	46.25 (5.75)	57.0 (11.53)	24.44 (3.25)	25.85 (11.48)
Gender	Male	53%	50%	50%	100%	67%	57%
	Female	47%	50%	50%	0%	33%	43%
Mean Years Teaching Experience (SD)		1.22 (2.33)	5.00 (5.66)	15.75 (6.30)	30.33 (17.50)	1.67 (1.58)	4.53 (8.59)
Recent Subject Area Taught	Astronomy	9%			33%		8%
	Biology	12%					8%
	Chemistry	12%		13%			9%
	Earth Science		50%			11%	3%
	Physics	46%		87%	33%	89%	55%
	Other	14%	50%		33%		11%
	Not reported	7%					6%
Race	Native American	2%					1%
	Asian	7%					5%
	Hispanic	7%		12.5%		11%	8%
	White	84%	100%	75%	100%	89%	85%
	Not Reported			12.5%			1%

Table 3.

Frequency of Scores for each item on each Dimension (N = 65)

	FA Dimension Score			SCI Dimension Score	
Item	0	1	2	0	1
1	15%	75%	10%	5%	95%
2	41%	54%	5%	79%	21%
3	25%	68%	7%	87%	13%
4	40%	53%	7%	52%	48%
5	36%	55%	9%	23&	77%