ZDM Mathematics Education
DOI 10.1007/s11858-012-0483-1

ORIGINAL ARTICLE

Classroom observations in theory and practice

Alan H. Schoenfeld

Accepted: 22 December 2012

Abstract This essay explores the dialectic between theorizing teachers' decision-making and producing a workable, theoretically grounded scheme for classroom observations. One would think that a comprehensive theory of decision-making would provide the bases for a classroom observation scheme. It turns out, however, that, although the theoretical and practical enterprise are in many ways overlapping, the theoretical underpinnings for the observation scheme are sufficiently different (narrower in some ways and broader in others) and the constraints of almost real-time implementation so strong that the resulting analytic scheme is in many ways radically different from the theoretical framing that gave rise to it. This essay characterizes and reflects on the evolution of the observational scheme. It provides details of some of the failed attempts along the way, in order to document the complexities of constructing such schemes. It is hoped that the final scheme provided will be of some value, both on theoretical and pragmatic grounds. Finally, the author reflects on the relationships between theoretical and applied research on teacher behavior, and the relevant research methods.

Keywords Teaching quality · Classroom observations · Coding scheme · Decision making · Rubric

A. H. Schoenfeld (✉)
Elizabeth and Edward Conner Professor of Education,
Education EMST, Tolman Hall #1670,
University of California, Berkeley,
CA 94720-1670, USA
e-mail: alans@berkeley.edu

1 Introduction and overview

1.1 Purposes of this paper

My first major purpose in writing this article is to lay out the complexities of constructing a classroom analysis scheme for empirical use, even when a general theory regarding teacher decision-making is available. On reflection, this complexity is inevitable: my work in problem solving (e.g., Schoenfeld, 1985, 1992) consisted of a decade of dialectic between evolving theoretical ideas and their empirical manifestations in problem solving courses, and my research on teacher decision making took nearly 20 years of theory building, intertwined with ongoing empirical studies. Capturing the dimensions of teaching in a manageable observation scheme is tremendously challenging, and readers rarely get to see the twists and turns of plausible but unworkable ideas that precede the presentation of the clean final product. I hope that revealing some of those pathways in this case will prove to be useful.

My second major purpose is to present the scheme itself—and with it, a new theoretical claim, that the dimensions highlighted within it may have the potential to be a necessary and sufficient set of dimensions for the analysis of effective classroom instruction. The dimensions are all well grounded in the literature, so there is some hope that this will turn out to be the case—although only time and more research will render that decision, as happened in the case of my problem solving book. Should the scheme prove viable as a classroom analysis tool, it may also have the potential to be used for charting teachers' professional growth and for coaching mathematics teachers.

My third major purpose, which I engage after the details of this analytic scheme and its development have been laid out, is to reflect on the multiple facets of performance reflected in different kinds of studies—those which engender and test theories of decision making, and those which examine decisions and actions with an eye toward how they shape learning. The same core constructs are involved, but they play out in different ways, and are most appropriately explored with different methods.

1.2 A framework for studying teacher decision making

The publication of my book *How We Think* (Schoenfeld, 2010) reflected the culmination of a decades-long research program into human decision-making. The book was aimed at providing a theoretical answer to the question, "what does one need to know in order to explain, on a moment-by-moment basis, the decisions made by an individual in the midst of a 'well practiced' activity such as teaching?" In theoretical terms, it argued that a characterization of the following four categories of the individual's knowledge and activity:

- resources (most centrally, knowledge)
- goals
- orientations (i.e., belief, values, preferences, etc.)
- decision-making (for routine decisions, as implemented by scripts, schemata, routines, etc.; for non-routine decisions, as modeled by a form of subjective expected utility)

is necessary and sufficient to enable one to construct a model of an individual's decision-making that is entirely consistent with the individual's behavior on a moment-by-moment basis. (That is, the decisions made by the model are in synch with those of the individual being modeled, on a line-by-line basis.) In methodological terms, the book provided a series of techniques for parsing and analyzing classroom activity structures:

- an iterated parsing of activities into nested sequences of phenomenologically coherent "episodes," reflecting cohesive sequences of classroom activity;
- the attribution of the teacher's relevant knowledge and resources, goals, and beliefs and orientations for each of these phenomenological episodes; and
- a description of the decision-making (either as part of a script, schema, or routine if things were going as planned, or a more complex analysis in the case of non-routine situations).

As my research group turned to conducting classroom analyses, it seemed reasonable to assume that both the major constructs in the theory and our methods of analysis would be central to the classroom analyses as well.

1.3 Ideas underlying the Algebra Teaching Study and Mathematics Assessment Project

The broad issue underlying the Algebra Teaching Study (US National Science Foundation grant DRL 0909815, Robert Floden and Alan Schoenfeld, Principal Investigators) and the Mathematics Assessment Project (Bill and Melinda Gates Foundation Grant OPP53342) is the relationship between classroom practices and the student understandings that result from those practices. Which classroom interactions, which pedagogies, result in students' "robust understanding" of important mathematics? Our expectation is that the theoretical frameworks that we develop for analyzing algebra classrooms will be applicable to the teaching of all mathematics content. In order for the scope of the work to be manageable, however, the Algebra Teaching Study chose to work on "contextually rich algebraic tasks"—not the stereotypical word problems of standard algebra texts, but problems that are stated in words and require some amount of analysis, modeling, and representation by algebraic symbolization in order to be solved. Such problems might be encountered in the eight or ninth grade in current US curricula. A sample task is given in Fig. 1. The overall scheme for our research is given in Fig. 2.

Hexagons
(Adapted from Mathematics Assessment Resource Service, http://www.noycefdn.org/resources.php, copyright 2003)

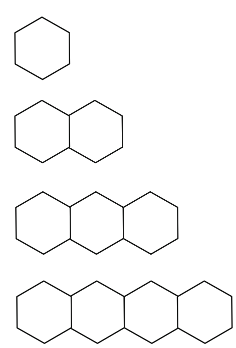
Maria has some hexagonal tiles. Each side of a tile measures 1 inch. She arranges the tiles in rows; then she finds the perimeter of each arrangement.

1 tile
Perimeter = 6 inches

2 tiles
Perimeter = 10 inches

3 tiles

4 tiles



(1) Find the perimeter of her arrangement of 4 tiles.

(2) What is the perimeter of a row of 10 tiles? How do you know this is the correct perimeter for 10 tiles?

(3) Write an equation for the perimeter p of a row of hexagonal tiles that works for any number of tiles, n, in the row. Explain how the parts of your equation relate to the hexagon patterns on the first page.

(4) Maria made a long row of hexagon tiles. She made a small mistake when counting the perimeter and got 71 inches for the perimeter. How many tiles do you think were in her row? Write an explanation that would convince Maria that her perimeter count is incorrect.

Fig. 1 A contextually rich algebraic task. (Adapted with permission from Mathematics Assessment Resource Service, http://www.noycefdn.org/resources.php, copyright 2003)

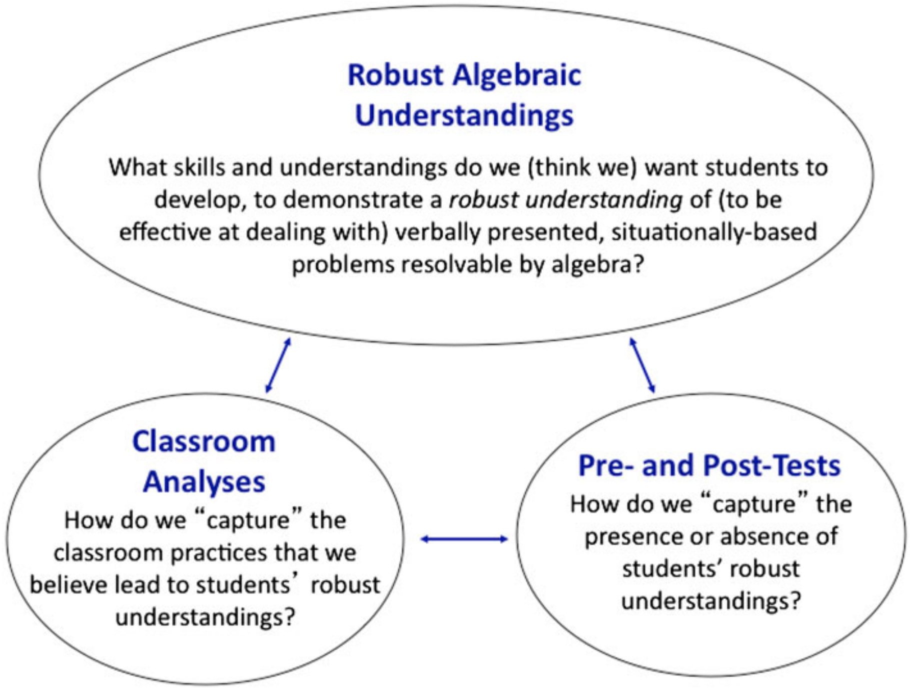


Fig. 2 The main issues addressed

The focus of the algebra part of our work is on "robust algebraic understandings"—on students' abilities to make sense of, and solve, contextually rich algebraic tasks (or more broadly, to engage in sense-making in algebra). Our goal is to explore the links between the two ovals at the bottom of Fig. 2: can we identify what we believe are productive classroom practices, and see if/how they are related to student performance? For pretests and posttests of algebraic performance, we selected a collection of contextual algebraic tasks from the Mathematics Assessment Resource Service, http://www.noycefdn.org/resources.php. Our challenge, then, was to develop a coding scheme for the "independent variable": could we craft a coding scheme that

(a) captures the aspects of teaching we believe are consequential for students' development of robust algebraic understandings, and
(b) is implementable in no more than, say, twice "real time?"

For the scheme to be workable on a large-scale basis, we wanted to be able to take notes on an hour-long lesson and then convert those notes into a set of scores on a coding sheet within another hour or so. Then, we would explore correlations between our codings and student performance on the pretests and post-tests. This kind of scheme, once robust, has a number of potential uses. A fundamental aim for the Gates Mathematics Assessment Project (MAP, 2012) is to trace teacher growth as teachers become increasingly adept at using the "formative assessment lessons" that MAP is building (see http://map.mathshell.org/materials/index.php). And, it may be that the analytic scheme presented at the end of this paper will—once there is evidence that teachers who score high on it do indeed have students who do well mathematically—provide a useful device for teacher coaching in mathematics.

For the balance of this paper, I focus on the creation of the analytic scheme and the issues that its creation raises.

2 Extant schemes

To sharpen our intuitions, the research group sought out videotapes of teachers recognized for their skill, and watched them at length. Then, over time, we looked at a wide range of schemes that other researchers or professional developers had constructed for the analysis of classroom interactions:

- Framework for Teaching (Danielson, 2011)
- Classroom Assessment Scoring System (Pianta, La Paro, & Hamre, 2008)
- Protocol for Language Arts Teaching Observations (Institute for Research on Policy Education and Practice, 2011)
- Mathematical Quality of Instruction (University of Michigan, 2006)
- UTeach Teacher Observation Protocol (Marder & Walkington, 2012)
- IQA, Instructional Quality Assessment, (Junker et al., 2004)
- PACT, the Performance Assessment for California Teachers (PACT Consortium, 2012)

- SCAN, the Systematic Classroom Analysis Notation (Beeby, Burkhardt, & Caddy, 1980)

Although each of these schemes had its virtues, each offered challenges with regard to our specific analytic goals. To be more explicit, we had at the time certain criteria that were tacit but that became more explicit as we worked on the scheme. Ultimately, we wanted a mechanism for capturing what takes place in mathematics classrooms that was (a) workable in roughly twice real time; (b) focused in clear ways on dimensions of classroom activities that were known in the literature to be important, (c) relatively comprehensive, in that the major categories of classroom actions noted in the literature were represented; (d) relatively comprehensible, in that the framework underlying the scheme cohered and was comprehensible; and, of course, that (e) the scheme had the requisite properties of reliability and validity. Although Fig. 2 can be interpreted in correlational terms (do high scores on classroom analyses correspond to high scores on student performance measures?), we hoped for more—that, ultimately, the (relatively few) dimensions of the analysis in the classroom analysis scheme would also, in the long run, provide a coherent and theoretically grounded basis for professional development.[1]

Here is a description of some of the challenges we faced in working with the schemes listed above[2]. Some, e.g., PLATO, did not focus on mathematics; none focused on assessment. Some, such as the Framework for Teaching, covered numerous teacher behaviors, at different levels of grain size; in looking at the rubrics we were unable to identify key constructs amidst the classroom activities coded. Some, such as the IQA, focused on one or more key constructs, such as classroom discourse, but they were too narrow for our purposes. We tried all of the schemes on tapes of what we perceived to be excellent teaching. Ultimately, none of the schemes jibed with our sense of what was central in good algebra teaching (that is, they did not meet the criteria given above). Things we saw the teachers doing, that we judged to be important, were not reflected in the coding we did.

3 First attempts: deriving a coding scheme from the research on decision making

As noted above, we had at our disposal an analytic framework that focused on key factors in the teacher's decision-making: the teacher's orientations (what does the teacher think is important about the content, about classroom interactions, about the students?), the teacher's goals for instruction, and the knowledge at the teacher's disposal for meeting those goals. We also had a mechanism, discussed above, for coding the lesson. The scheme had been used for research purposes, where we had the luxury of taking months to come to certainty about the codings we assigned. But, the classes we had coded for research purposes were extraordinarily complex. In contrast, most classroom instruction is not nearly as complex—and the goal of the current research was to do a quick parsing that met the standards of inter-rater reliability rather than trying to get every detail right. So, we tried to adapt the coding scheme discussed above.

The attempt was disastrous. It was easy to parse lessons into episodes and sub-episodes—for the most part, break points in classroom activity structures are easy to observe. But, the scheme had two fatal flaws. First, it called for a great deal of inference and/or interviewing on our part, in order to develop an understanding of the teacher's goals and orientations. Second, it was too teacher-focused—it did not capture the students' experiences adequately. For example, Phil Daro, one of the members of the ATS advisory board, has said that the most important predictor of student learning may be that the number of times that students get to say a second sentence in a row. (See also Franke, Kazemi, & Battey, 2007; Franke & Webb, 2010; Franke, Webb, Chan, Ing, Freund, & Battey, 2009.) This kind of consideration was absent from the decision-making scheme. We decided to abandon the research scheme as a viable method for the relatively rapid coding of classroom activities that we desired. Ultimately, as described in what follows, various aspects of the research scheme—e.g., the parsing of a lesson into episodes, and the documentation of the results of their in-the-moment decision making (grounded in their orientations, beliefs, and goals) became parts of our current coding system. But, the need to focus on activity structures for all of the classroom participants, and to not engage in deep and extended analyses of what the teachers knew, believed, and were trying to achieve, mandated very significant changes in approach.

4 Second attempt: a potentially comprehensive framework

The research group turned to a more straightforward analysis. The idea was simple in outline. Consider a matrix in which the columns represent desired student outcomes, and the rows represent important aspects or types of classroom interactions. We had three major student outcomes, listed as follows:

[1] A large study funded by the Gates Foundation, the Measures of effective Teaching (MET) project (2012), did examine correlations between student learning and performance on some of the measures above.

[2] This is not the place to provide an extensive critique of the extant schemes, or a comparison of them. Such a critique will be provided in (Algebra Teaching Study, 2013, in preparation).

A. Access. How much "room" was there for all students to engage mathematically?
B. Accountability. In what ways were students held to high mathematical standards?
C. Productive dispositions. Did students develop appropriately productive mathematical dispositions and habits of mind?

We identified four central points of focus for our classroom analyses:

1. The mathematics
2. Opportunities for mathematics learning
3. The classroom community
4. The individual learner

This structure produced a straightforward summary matrix for characterizing the learning environment. See Fig. 3.

The approach in Fig. 3 offered two main challenges. First, the underlying analytic superstructure was quite complex. Each of the cells in the matrix is a summary cell—and the details required to assign a summary score for that cell were anything but simple. Each of the cells in Fig. 3 had a number of contributory sub-dimensions; see Fig. 4.

Second, we had a series of observational codes that contributed to scores. There were codes for teacher, students, and task. For example, one of the 12 teacher codes was "Teacher pushes for conceptual understanding"; one of the student codes was "Students question and evaluate mathematical ideas, whether they come from the teacher or from classmates"; and one of the task codes was "Task requires students to justify, conjecture, interpret." A score on any of these codes could contribute to numerous scores in the three-by-four matrices in Figs. 3 and 4.

This scheme, while highlighting many things we thought were important, was very unwieldy. Despite the seeming simplicity of Fig. 3, the list of codes was somewhat ad hoc and the actual mechanics of coding lessons almost impossible.

5 Subsequent attempts: tries at simplicity, interwoven with evolving complexity

For nearly 2 years the research group tried, in various ways, to move the scheme forward and to make it workable. Until we arrive at the penultimate scheme, extensive detail is not important. My purpose here is to highlight the challenges of doing such work, and the many ways in which good ideas turn out to be difficult to implement. Illustrative detail is given where warranted.

5.1 Levels of mathematical activity

In reviewing extant schemes, we noted that some focused, either in whole or in part, on general patterns of classroom activity; some focused on mathematical activity. We tried a 3-level analytic scheme: general activity (how well organized and managed is the classroom, how interactive; how often do students get to speak, and in what ways?); mathematical activity (what are the sociomathematical norms in the classroom; what are the standards of explanation?) and specific algebraic activity (what supports are there for making sense of complex contextual word problems?) This proved very hard to organize and manage; we had three simultaneous coding schemes at the three levels of activity description.

	Access (what the teacher gives/allows)	Accountability (what the teacher expects/demands)	Productive Dispositions (what the teacher receives from students)
Strand	Dimensions (codes)	Dimensions (codes)	Dimensions (codes)
Mathematics	Students are able to experience the vibrancy and power of the domain of mathematics	Mathematical exploration and discussion should be accurate. Reasoning and justification should be tied to mathematics.	Students construct mathematics, attempting to discover rather than just receive.
Mathematics Learning	Students are given a chance to learn mathematics. This requires making making mathematics learning practices explicit and accessible.	Students are expected to engage productively in the mathematics learning process, sustain efforts, and contribute to finding solutions.	Students are interested in learning mathematics.
Classroom Community	No students are marginalized in the classroom community. All students have a chance to engage and participate.	Students have an obligation to their teacher and peers to be respectful and helpful. Students are not just participants but leaders of the classroom community.	Students contribute and participate as a community of mathematics practicioners.
Individual Learner	The classroom respects the uniqueness of each individual student, and gives appropriate affordances.	Students have an obligation to themselves to learn mathematics, and productively engage the subject matter.	Students sustain efforts as learners. Students take risks and believe that they can succeed.

Fig. 3 Central features of our second attempt

	Access	Accountability	Productive Dispositions
Strand	Dimensions (codes)	Dimensions (codes)	Dimensions (codes)
Mathematics	a) the teacher presents tasks in a way that demand rich mathematical engagement b) tasks provide opportunities to engage higher-level mathematical thinking	a) teacher presses for accuracy b) teacher carefully and accurately presents mathematical ideas c) multiple representations are required, used, and connected by teacher, students, and task d) teacher and students use academic language c) discussion among students is math-focused	a) students construct mathematics rather than wait to receive it b) students generate/explain ideas c) students question, challenge, evaluate ideas
Mathematics Learning	a) teacher is explicit about what to do on a given problem b) teacher is explicit about how to use formal math language c) teacher is explicit about how to reason mathematically d) students facilitate discussions e) students manage logistics f) students set the agenda/have choice in activities	a) teacher expects students to be able to learn mathematics b) teacher expect students to persist in mathematics learning c) teacher asks probing questions/elicits reasoning and justification e) teacher checks for understanding and provides feedback during instruction	a) students are excited, curious, or interested to engage math b) students seek multiple solutions to a single problem c) students don't just seek solutions but to understand why they work
Classroom Community	a) teacher provides feedback b) teacher relates and connects student ideas to one another c) teacher revoices/marks student contributions d) teacher positions students as equals e) students give and receive feedback from other students	a) authority is distributed between students and teacher b) authority is distributed between existing and new ideas c) students question and evaluate each other and teacher	a) students work collaboratively b) students respect one another's ideas c) students accept feedback from other students/teacher d) students acknowledge others' contributions
Individual Learner	a) teacher permits use of non-dominant language b) teacher provides students time to work independently c) teacher builds on students' prior knowledge, connects mathematical ideas d) students engage the mathematics on their own level e) tasks have multiple entry points f) problem contexts respect students' cultural backgrounds/prior knowledge	a) students have a role as mathematical authorities b) students participate in classroom activities	a) teacher positions students as competent b) teacher positions students as *capable* of doing the math - from Ball's MQI and Cohen's complex instruction c) students take risks a) students work hard b) students sustain efforts to reach learning goals (they don't give up after 2 minutes)

Fig. 4 Sub-dimensions of our second attempt

5.2 Activity Structures

In an attempt to rein in complexity, we returned to the idea of "episodes," periods of time during which the class is engaged in one relatively coherent type of classroom activity. This time, when we coded, we would parse a lesson into episodes, classify the type of episode, and then ask relevant questions about each episode. The activity types were:

- Task introduction
- Mathematical discussion
- Small group work
- Independent student work
- Post-Lesson analysis

For each of these activity structures we had codes for relevant activities. Figure 5 provides the codings for teacher and student behavior during mathematical discussions.

This version of the scheme, although more easy to code chronologically than earlier versions (we could take notes and identify episodes, then code behaviors within episodes) was still problematic. It had a large number of codes, which required simultaneous coding (e.g., for one classroom discussion, every one of the seven teacher behaviors and three student behaviors needed to be coded). It produced a series of coding values for different types of interactions, but there was no clear theoretical rationale for combining those numbers. Assigning some code values required a large degree of inference and value judgment. Consider for example teacher behaviors 4 and 5, assessing whole class understanding and pacing class discussions. Much ongoing teacher assessment is unspoken. Thus, it may be difficult if not impossible to know to what degree a teacher is assessing student understanding and modifying the pace of class in response to what he or she sees in student work or hears students say. And, how does one know whether the pacing or the examples are "appropriate" for most students? Some cases may be clear, but some may be subtle; some may depend on a teacher's goals or style, but be effective. Thus, although this version had some desirable elements, it was not yet workable.

5.3 Attempting to use the didactic triangle to provide structure

As the number of codes had increased, the scheme became increasingly unworkable. The idea of activity structures made sense, but coding multiple dimensions within any activity structure was a challenge. Thus we moved toward more fine-grained activity structures, with the expectation that coding within each activity structure would be more straightforward. At one point we had fifteen activity structures of relevance, some of which were as follows:

- Teacher leads whole class discussion
- Teacher prepares students for a new task
- Students ask a mathematical question
- Navigating a task's language or context
- Summarizing the math in a task

Mathematical Discussion (MD)				
		Level of Emphasis		
	Description	Low: 1	Average: 3	High: 5
Teacher Behavior				
1	Richness of Mathematics	If underlying mathematics concepts are engaged, the engagement is superficial.	Underlying mathematics concepts are engaged, but not in ways that make connections to other mathematical ideas.	Underlying concepts are central to the discussion. The emphasis is on understanding why and making connections between mathematical ideas.
2	Teacher's Mathematical Integrity	Teacher's mathematics contains significant errors.	Teacher's mathematics is generally correct but does not help students focus on key ideas.	Teacher's mathematics is generally correct and helps students focus on key ideas.
3	Soliciting Student Reasoning	Teacher does not solicit student ideas, or only asks for answers, not reasoning or justification.	Teacher asks students to provide some reasoning and explanation about mathematical ideas, but student participation is mostly limited to student-teacher interactions.	Teacher presses students for reasoning and justification of ideas/solutions, building the discussion using student ideas, and pressing students to question/analyze each other's reasoning.
4	Assessing Understanding (Whole Class)	Teacher does not assess student understanding or only does so superficially.	Teacher makes some attempt to check whether students are following key ideas of the discussion, but fails to productively use that information.	Teacher makes sure students are following the discussion and assesses their understanding of important mathematical ideas (by using student work and asking questions). The flow of the lesson/discussion is modified as appropriate based on these assessments.
5	Pacing of Discussion	Teacher provides an excessive amount of time or an insufficient amount of time for students to engage with questions/concepts (e.g. teacher answers own questions or always calls on first hand).	The pace of the discussion is engaging/accessible for most students, but the teacher spends too little time on some important topics or too much time on less important topics.	The pace of the discussion is engaging/accessible for most students.
6	Opportunities for Deeper Mathematical Conversations	Teacher misses opportunities for deeper mathematical conversations.	Teacher leverages opportunities for deeper, conceptual conversations, but often resolves the mathematics for students.	Teacher opens deeper, conceptual conversations, and persists in having students' resolve mathematical questions as much as possible.
7	Addressing/Engaging Misconceptions	Teacher leaves misconceptions unaddressed except when they are treated as "wrong answers" and corrected.	Teacher addresses some misconceptions but either (a) major misconceptions are left unaddressed or (b) the "fixes" are somewhat superficial.	Teacher engages misconceptions, probing for misunderstandings and building on partial understandings.
Student Behavior				
1	Participation	There is little student participation.	Participation is limited to a subgroup of students.	Many students participate.
2	Risks	Students don't share ideas.	Students share ideas when they are mostly certain they are correct	Students take risks in sharing their ideas
3	Student Explanations	Students don't explain their ideas or solution processes.	Students' explanations consist of what they did/think but not why.	Students explain why their solutions or ideas work, as appropriate.

Fig. 5 Mathematical discussions coding detail

For each of these different activity structures, we asked three sets of questions drawn from the didactic triangle: What can we say about the relationships between the teacher and students, between teacher and the mathematics, and the students and the mathematics? But sometimes other considerations were relevant, for example how well the task supported multiple representations or student argumentation. The result was a large matrix that turned out to be only semi-coherent—see Fig. 6 for the first four activity structures (situations) examined in the scheme.

This time the semi-coherence turned out to be productive. It was clear that almost everything we thought was of importance was somewhere in the giant fifteen-by-six matrix of which Fig. 6 is a part. But, with the fifteen activity structures represented in rows A through O, and between 2 and 6 aspects of the lesson coded for each row, something had to be done. We had reached completeness of coverage; but we had lost comprehensibility. The challenge was then to distill the content in the matrix, in ways that cohered logically and that fit with the literature.

The next step was simple, in concept. What if one took each non-empty cell in the matrix and asked: What fundamental issue from the literature does this cell address? The idea was to cluster similar cells—to create what are, in essence, mathematical equivalence classes—and to identify those equivalence classes as the fundamental dimensions of analysis. Consider row A of Fig. 7, for example. The first cell, "deciding who gets called on," is fundamentally about *equity and access*. In an equitable class, all students have the opportunity to contribute, and the teacher has a range of mechanisms for encouraging and supporting such contributions. The second and third cells are concerned with student *agency and authority*. A major issue is, when if ever do students get to develop a mathematical voice? That is, when do they get to propose ideas and answers, defend them, and become recognized as producers of mathematics themselves? *Equity and access* and *agency and authority* are two of the fundamental dimensions that emerged from our analyses. Broadly speaking, the goal was to classify each cell in the matrix as belonging to one of a relatively small number of categories that (a) had internal coherence, (b) represented an important vector in the literature, and (c) could be clearly distinguished from the others. These categories would become the dimensions for analysis.

6 The current version of the TRU Math (teaching for robust understanding of mathematics) scheme

In this section I present our current analytic scheme and the rationale in the literature for it. As noted above, the origins of this version of the scheme lay in looking for equivalence classes of important classroom activities. As I undertook

#	Situation	Teacher - Student	Student - Content	Content - Teacher	Other
A	Teacher Leads Whole Class Discussion	*Deciding Who Gets Called On*	*Nature of Student's Response*	*Nature of Teacher's Solicitations*	
		1 Only the first student that raises his/her hand is the one that gets called on.	1 Most of the student responses during this chunk consisted of only single-word responses (e.g., IRE).	1 Most of the teacher's solicitations during this chunk consisted of IRE-style questioning.	1
		2 Beyond the first student, at least one other student who raised his/her hand gets called on to respond to a given question.	2 Most of the student responses during this chunk consisted of procedures (possibly including an answer) for how to solve a problem.	2 Most of the teacher's solicitations during this chunk were of a procedural nature (i.e., how students solved a problem).	2
		3 Teacher uses techniques to actively engage students who do not volunteer (e.g., wait time, popsicle sticks, cold calling).	3 Most of the student responses during this chunk consisted of extended explanations (multiple sentences with a reason).	3 Most of the teacher's solicitations during this chunk asked students to explain why their answer/procedure makes sense.	3
B	Teacher Preps Students for a New Task	*Setting Process Expectations*	*Students' Opportunity to Engage the Task*	* Setting Product Expectations*	*Encouraging Multiple Solution Paths*
		1 Teacher tells students to get started without setting process expectations.	1 Teacher solves the task for students before they have a chance to engage it.	1 Teacher tells students to get started without setting product expectations.	1 The task/introduction strongly suggests a single solution path.
		2 Teacher sets process expectations (e.g., amount of time for task, how students should organize themselves).	2 Students are handed the task and are told to get started with no teacher intervention.	2 Teacher sets expectations about final product (e.g., by providing a scoring rubric, showing examples of high quality work).	2 The task/introduction affords multiple potential solution paths.
		3 Teacher engages students in mutually setting process expectations.	3 Teacher engages students in brainstorming possible approaches to the task without explicitly showing them how to do it.	3 Teacher engages students in mutually setting expectations for final product.	3 The task/introduction encourages/requires multiple solution paths and/or the contrast of different solutions.
C	Student Asks a Mathematical Question	* Whether/How Teacher Takes up the Student's Question *	* Cognitive Demand of Student's Question *	N/A	
		1 Teacher ignores or dismisses the question.	1 The student asks about whether an answer is correct or not (i.e., a "WHAT" question) or a non-specific question (e.g., "I don't know how to get started!")	1	1
		2 Teacher gives an explanation directly answering the student's question.	2 The student asks a specific question about HOW to do a procedure.	2	2
		3 Teacher engages the student/class in answering the question (e.g., acting as a guide).	3 The student asks WHY something works.	3	3
D	Navigating a Task's Language or Context	*Who Does the Navigating*	N/A	*Opportunity to Engage Task as Written*	
		1 Teacher does not check if students are comfortable with the language or problem context.	1	1 Teacher boils the language or context out of the task, leaving only the quantities needed to get an answer.	1
		2 Teacher checks if students are comfortabel with the language or problem context, but does the work for the kids (e.g., defines unfamiliar words for students, paraphrases the problem context).	2	2 N/A	2
		3 Students and teachers work collaboratively to build students' understanding of the language or context in the problem.	3	3 Teacher gives students an opportunity to work with the task as stated.	3

Fig. 6 Part of the didactic frame

this work in earnest, I came to a deeper understanding of the kind of structure that I was seeking. The following analogy may be helpful for understanding what the framework embodies.

The product of my research on problem solving (Schoenfeld, 1985, 1992) was a framework for the analysis of the success or failure of problem solving attempts. The framework focused on four categories of behavior:

- the knowledge base
- problem solving strategies
- metacognition, specifically monitoring and self-regulation
- belief systems, and the practices that gave rise to them.

Perhaps most important, I claimed that the four categories were both necessary and sufficient for the analysis of problem solving attempts, in the following sense. They were necessary in that one had to consider all of them when evaluating a problem solving attempt—the cause of success or failure might reside within the knowledge base, access to strategies, metacognition or beliefs, and one might miss the cause unless all were examined. They were sufficient in the sense that the cause of success or failure would reside in one of those categories; no other dimensions of problem solving need be examined. In addition, each of the categories cohered, and there was relatively little overlap between categories.[3]

[3] It is impossible to separate the categories completely – a strategy is part of one's knowledge base, for example, and some metacognitive acts are strategic. However, there are better and worse decompositions. The idea is to aim for a "nearly decomposable system," a decomposition in which the parts cohere internally and have minimal overlap. One might, for example, divide the human body into a series of parts: arms, legs, torso, head – but that makes no sense physiologically, in terms of function. On the other hand, a decomposition into respiratory system, circulatory system, muscular system, skeletal system, and so on, does make sense. The systems themselves cohere, and, although there is overlap and interaction, e.g., between the circulatory and respiratory systems, it makes sense to talk of them (almost) independently.

Level	Mathematical Focus, Coherence and Accuracy	Cognitive Demand	Access	Agency: Authority and Accountability	Uses of Assessment
1	Classroom activities are purely rote, OR disconnected or unfocused, OR consequential mistakes are left unaddressed.	Classroom activities are structured so that students mostly apply familiar procedures or memorized facts.	Classroom management is problematic to the point where the lesson is disrupted, OR a significant number of students appear disengaged and there are no overt mechanisms to support engagement.	The teacher initiates conversations. Students' speech turns are short (one sentence or less) and shaped or constrained by what the teacher says or does.	The teacher may note student answers or work, but student reasoning is not surfaced or pursued. Teacher actions are limited to corrective feedback or encouragement.
2	The mathematics discussed is relatively clear and correct, BUT connections between procedures, concepts and contexts (where appropriate) are either cursory or lacking.	Classroom activities offer possibilities of conceptual richness or problem solving challenge, but teaching interactions tend to "scaffold away" the challenges and mostly limit students to providing short responses to teacher prompts.	The class is engaged in mathematical activity, but there is uneven participation and the teacher does not provide structured support for many students to participate in meaningful ways.	Students have a chance to say or explain things, but "the student proposes, the teacher disposes": in class discussions, student ideas are not explored or built upon.	The teacher refers to student thinking, perhaps even to common mistakes, but specific student ideas are not built on (when potentially valuable) or used to address challenges (when problematic).
3	The mathematics discussed is relatively clear and correct, AND connections between procedures, concepts and contexts (where appropriate) are addressed and explained.	The teacher's hints or scaffolds support students in "productive struggle" in building understandings and engaging in mathematical practices.	The teacher actively supports (and to some degree achieves) broad and meaningful participation, OR what appear to be established participation structures result in such participation.	Students put forth and defend their ideas. The teacher may ascribe ownership for students' ideas in exposition, AND/OR students respond to and build on each others' ideas.	The teacher solicits student thinking and subsequent instruction responds to those ideas, by building on productive beginnings or addressing emerging misunderstandings.

Episode Type		Mathematical Focus, Coherence and Accuracy	Cognitive Demand	Access	Agency: Authority and Accountability	Uses of Assessment
W	**W**hole Class Activities	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric
G	Small **G**roup work	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric
P	Student presentations	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric	scored on a 3-point rubric
I	**I**ndividual work	not scored	not scored	not scored	not scored	not scored

Fig. 7 Outline of the TRU Math scheme

My goal for the TRU Math scheme is for the equivalence classes that emerge from the analyses described in the previous section to have similar properties—that there would be a relatively small number of categories of classroom activities for analysis; that they would be necessary, in that to ignore any of them would run the risk of missing an essential component of instruction; and that they would be sufficient, in that no other categories would be necessary for analysis. Although it is too early in the process to be confident, I think that there is a good chance that the dimensions discussed below have those properties. I begin by introducing the five dimensions. These are the columns of the TRU Math scheme, which is in the form of a matrix. Having described the fundamental dimensions of the scheme, I provide an outline of the analytic structure of the matrix. The four rows of the matrix are a range of classroom activity structures. The "basic" matrix, of the form {activity structures} × {dimensions}, provides the core analytic structure of our general approach. As discussed below, this core suffices as a general classroom tool; but more detail is needed for close examination of any particular topic or dimension (e.g., the specifics of algebra learning, or assessment).

6.1 Dimensions of TRU Math

The dimensions are as follows.

1. *Mathematical Focus, Coherence and Accuracy*. To what extent is the mathematics discussed clear, correct, and well justified (tied to conceptual underpinnings)?
2. *Cognitive Demand.* To what extent do classroom interactions create and maintain an environment of intellectual challenge?
3. *Access.* To what extent do classroom activity structures invite and support active engagement from the diverse range of students in the classroom?
4. *Agency, Authority and Accountability.* To what extent do students have the opportunity to make mathematical conjectures, explanations and arguments, developing "voice" (agency and authority) while adhering to mathematical norms (accountability)?
5. *Uses of Assessment.* To what extent is student reasoning elicited, challenged, and refined?

Dimension 1: Mathematical Focus, Coherence and Accuracy. This dimension pertains to the richness and centrality of the mathematics as it plays out in the classroom. In the US context, there is a history of major curriculum documents. The US National Council of Teachers of Mathematics issued two sets of *Standards,* in (1989) and (2000). These were voluntary standards, in the sense that each of the 50 individual states in the United States was free to adopt its own standards and assessments—and did. More recently, the Common Core State Standards Initiative issued a consensus set of mathematics Standards (CCSSI-M, 2010), which have been adopted by 45 states. A major feature of CCSSI-M is a focus on mathematical *practices,* for example making sense of problems and persevering in solving them, reasoning abstractly and quantitatively, constructing and critiquing viable arguments, modeling with mathematics, and using appropriate tools strategically. Scores along the mathematics dimension reflect the opportunities for students to engage with important mathematical content and practices, in a way that is focused and coherent, tied to conceptual underpinnings (in contrast, for example, to the rote memorization of procedures).

Dimension 2: Cognitive Demand. In a series of major articles, Stein, Henningsen, and colleagues (Henningsen & Stein, 1997; Stein, Engle, Smith, Hughes, 2008; Stein, Grover, & Henningsen, 1996) explored the role of classroom discourse in either maintaining or diluting the mathematical richness of tasks with which students engage. Henningsen & Stein (1997) document five factors that appear to be "prime influences associated with maintaining student engagement at the level of doing mathematics:" mathematically rich tasks, "teacher scaffolding that enables students to grapple with the task without sacrificing or diluting the important mathematics in it," adequate time, modeling of high quality performance, and a "sustained press for explanation and meaning." Henningsen and Stein (1997) note three major types of decline from powerful engagement as well: (1) decline into using procedures without connection to concepts, meaning, and understanding, (2) decline into unsystematic exploration and lack of sustained progress in developing meaning or understanding, and (3) decline into activities with little or no mathematical substance. Scores along this dimension reflect whether the mathematics has been "proceduralized" to the point where there is little true mathematical engagement, or whether students get to engage in "productive struggle" as they work on the mathematics.

Dimension 3: Access and Equity. Access to powerful and meaningful mathematics is important for all students (Moses, 2001; Schoenfeld, 2002). There is a long history of differential achievement in mathematics by students from varied racial, ethnic, and economic backgrounds (Secada, 1992), which, it has been argued, can be tied to differential access to opportunities to learn (Oakes, Joseph, & Muir, 2001). While one obvious source of this differential access is tracking, which is outside of the scope of a classroom observation scheme, another is the pattern of discourse within classrooms. Who has the opportunity to engage with mathematics in ways that are likely to lead to learning? Do all students have opportunities to discuss mathematical ideas with some frequency (American Association of University Women, 1992)? Are there multiple opportunities to develop and display competence for each student (Cohen 1994), and for students to build understanding based on the knowledge they bring with them into the classroom (González, Andrade, Civil, & Moll, 2001; Zevenbergen, 2000)? This dimension of our observation scheme attempts to address these questions to the extent that it is possible to do so in discrete classroom observations.

Dimension 4: Agency, Authority and Accountability. Mathematics learning is active, not passive. In a productive learning environment students have the opportunity to see themselves as doers of mathematics—to develop a sense of agency—and to act accordingly (Engle, 2011; Engle & Conant, 2002; Schoenfeld, in press). Agency, of course, is part of one's mathematical identity and disposition. The roots of "authority" reside in the word "author": the idea is that students create, or author, mathematical ideas and their justifications (thus becoming authorities). At the same time, students are not free to invent without constraint: they make conjectures, but are then multiply accountable—to the discipline, to the teacher, and to other students.

The discourse structures supported by the teacher can foster or inhibit agency, authority, and accountability. The Institute for Research on Learning (IRL, 2011) argues that the norms of "Accountable Talk" play facilitative roles in developing student agency and authority. The following teacher moves can be productive: revoicing (the teacher restates something a student said, attributing it to the student), asking students to restate someone else's reasoning, asking students to apply their own reasoning to someone else's reasoning, prompting students for further participation, asking students to explicate their reasoning, and challenging students' reasoning by asking for counterexamples, etc. (Resnick, O'Connor, and Michaels, 2007). Scores along this dimension reflect whether the classroom environment provides students with opportunities to develop agency and authority, subject to the appropriate mathematical norms (accountability).

Dimension 5: Uses of Assessment. In contrast to commonplace practices of classroom assessment being separate from instruction and serving a predominately evaluative function (Shepard, 2000), major policy documents in math education research assert that assessment should become an integral component of instruction (NCTM, 1995; NRC, 2001; NRC, 2005). Black and Wiliam's (1998) widely cited review of the research literature on formative assessment documents substantial learning gains that result from teachers' use of formative assessment practices. When assessment becomes an integral and ongoing part of the learning process, as opposed to an interruption of classroom activities, students' thinking takes on a more central role in determining the direction and shape of classroom activities (Shepard, 2000; Shafer & Romberg, 1999; de Lange, 1999). In consequence, teachers' instruction can more adeptly support and enhance students' individual and collective reasoning (Webb & Romberg, 1992). Additionally, through self- and peer-assessments, students can be positioned through the construction of particular classroom norms to become more reflective regarding their own learning processes (Shepard, 2000). As noted in the introduction to this article, the Mathematics Assessment Project (http://map.mathshell.org/materials/index.php) is constructing 100 formative assessment lessons, whose goal it is to support teachers in their ability to elicit, challenge, and refine student thinking. Scores along this dimension reflect the degree to which assessment is used productively in the classroom.

6.2 Activity Structures in TRU Math

As in earlier versions of our scheme, the horizontal rows of our coding matrix represent classroom activity structures. The idea behind coding a lesson is that the lesson is parsed sequentially into a series of episodes or activity structures that are relatively short (less than five minutes) and phenomenologically coherent. In constructing the set of activity structures, we strove for the following properties. We wanted the list to be relatively short, but to contain the activities that are likely to be consequential in terms of the five dimensions. If possible, we want the list of activity structures to be disjoint (no more than one activity structure per episode), so that each episode only needs to be coded once. Although this list is still subject to revision, our current set of activity structures is the following:

- Whole Class Activities, including as subsets Topic Launch, Teacher Exposition, and Whole Class Discussion;
- Small Group Work; and
- Student Presentations.

We also note, but do not typically code, periods of

- Individual Student Work (which is noted, but not typically coded)

For each of the first three activity structures, (Whole Class, Small Group, and Student Presentations) the classroom episode is rated on a scale of 1–3 for each dimension. Ultimately, the scores are aggregated over the lesson.[4]

[4] Episodes are between 45 s and 5 min. Our scoring guide provides rules for carving longer periods of activity (say, 15 min of whole class discussion) into episodes that are no longer than 5 min. Currently we are exploring a number of different ways of aggregating data across episodes.

Figure 7 presents the outline of the current version of the core scheme. In the full version, there are rubrics for assigning scores of 1, 2, or 3 for each episode type, for each dimension. (That is, there are 20 separate 3-point rubrics for assigning scores to each cell in the 4-by-5 matrix.) The top part of Fig. 7 represents the summary rubrics for each dimension, and the bottom part shows the dimensions of the 4-by-5 matrix.

This, of course, is a skeletal scoring summary—there is a substantial amount of supporting detail. As seen in the bottom section of Fig. 7, each of the cells representing whole class activities, small group work, and student presentations is scored on a 3-point rubric.[5] To pick just one as an example, the student presentations rubric under "uses of assessment" (row 3, column 5) is scored as follows. A score of 1 indicates that "when errors are made, teacher does not engage presenter or other students in discussion; OR, actions are simply corrective." A score of 2 indicates that "Teacher probes presenter/class for reasoning and uses this to elaborate on correct ways to do the mathematics." A score of 3 indicates "Teacher comments and questions support presenter and other students in airing and vetting the ideas behind the work they produce." These statements are in themselves brief summaries of the gist of an episode. We are compiling an extensive scoring guide, which provides illustrative examples to indicate what scores are given under what circumstances.

I referred to the 4-by-5 matrix above as the "core scheme," in that it is mathematically general. As given, the 4-by-5 matrix provides general detail that, we think, will correlate well with student outcomes in any mathematics course. However, a more fine-grained lens is necessary for analysis in any particular mathematics content area. As noted above, our current work is in algebra. Hence a part of the scheme is focused on teaching for robust understanding of algebra word problems. In addition to dimensions 1 though 5, classroom activities for the algebra work are also coded for how well instruction supports:

- Reading and interpreting text, and understanding the contexts described in problem statements.
- Identifying salient quantities in a problem and articulating relationships between them
- Generating representations of relationships between quantities
- Interpreting and making connections between representations
- Executing calculations and procedures with precision
- Checking plausibility of results
- Opportunities for Student Explanations
- Teacher instruction about Explanations
- Student Explanations and Justifications

This addition of an "algebra word problem module" makes the scheme algebra word problem specific. One could easily replace this algebra module with one for geometry, or calculus, or other content. Similarly, researchers with specific interests in cognitive demand, access, agency, and assessment could expand the scheme (either by adding rubrics, as above, or by specifying more activity structures) to flesh out the scheme to the desired level of detail. Thus, for a close look at assessment one might delineate as separate episodes segments of student work where students are putting together posters for presentation, demonstrating their current work on mini-whiteboards, etc.

7 Discussion

As noted in the introduction, my first major purpose in writing this chapter was to lay out the complexities of constructing a classroom analysis scheme for empirical use, even when a general theory regarding teacher decision-making is available. It took my research group 3 years of concentrated effort to create the analytical scheme summarized in Fig. 7, even though we had at our disposal a robust analytic framework for characterizing teacher decision making. I have summarized some of the twists and turns in the development of the scheme, because I think it is important to do so. In the literature we often find polished gems, whose contorted history has been obscured. There is, I believe, much to be gained from examining the ways in which our understandings develop.

My second major purpose was to present the scheme itself. All I have at this point, as I had more than 25 years ago with regard to problem solving, is an intuitive sense that the dimensions highlighted in the scheme have the potential to be necessary and sufficient for the analysis of effective classroom instruction. There is no doubt about their importance, in general: each of the dimensions has a solid grounding in the literature. How important they will turn out to be, individually or in combination, remains to be seen. If they do hold up analytically, then there is a next set of challenges. On the one hand, my work on teacher decision making indicates that teachers' resources (especially knowledge), goals, and orientations (especially beliefs about students and mathematics) are highly consequential. On the other hand, this work suggests the dimensions of powerful classroom environments. The challenge for professional development thus becomes, how can we create contexts for professional growth, in which teachers' knowledge and resources, goals, and orientations can evolve productively in ways that enable the teachers to

[5] In most cases, the rubrics for different episodes are different, taking into account the specifics of that kind of episode. In a small number of cases, the rubrics for a particular dimension are identical.

craft instructional environments that score well on the dimensions of the scheme indicated in Fig. 7?

Third, the two theoretical frameworks described in the previous paragraph and our intentions for professional development provide an opportunity for me to reflect on the roles of various theoretical constructs and research methods in exploring productive classroom behavior. Here I offer three observations:

1. Depending on one's focus, different constructs may appear to play more or less central roles.
2. No matter what the claim concerning teaching and learning, a dialectic with empirical observations is essential.
3. Getting at "what counts" requires multiple lenses, methods, and perspectives.

7.1 Observation 1: the varying salience of fundamental constructs

Consider the three main constructs in the theory of teachers' decision making (Schoenfeld, 2010): knowledge and resources, goals, and orientations. The goal of that body of research was to explain teachers' in-the-moment decision making—to be able to explain, on theoretical grounds, how and why teachers made each decision they did while in the midst of teaching. This theoretical framing required a micro-analytic approach, the questions being, what knowledge does the teacher have potentially at his or her disposal, and for what reasons does he or she make particular choices? These three constructs—knowledge and resources, goals, and orientations—play more of a background role in the research on powerful classroom environments that has been the focus of this article. The central question for this kind of classroom research is, what are the key dimensions of the learning environment, as experienced by the students? It goes without saying that the teacher's decision making plays a fundamental role in *shaping* the environment: a teacher cannot teach content or use pedagogical techniques of which he or she is unaware, and how much of a priority the teacher assigns to (for example) mathematical sense making or giving all students an opportunity to participate meaningfully in classroom activities is vitally important. Yet, what matters in the classroom are the activity structures as the students experience them. Hence, the teacher's knowledge, goals, and orientations are "backgrounded" in this context, as classroom activity structures are highlighted.

Interestingly, these constructs are likely to be foregrounded once again when one turns to professional development. The goals of professional development are to enhance the learning environment, but the means of achieving that improvement lie in the enhancement of the teacher's capacity to craft a more powerful learning environment. Teacher knowledge is obviously important, as are material resources; a teacher cannot implement what he or she does not know or does not have the resources for. However, effective professional development will also have to target teachers' beliefs and goals. If these remain unchanged, new knowledge may not be put to use.

7.2 Observation 2: the need for a dialectic between theory and empirical observation

I am convinced that neither theoretical nor empirical research can thrive without the other. Thus, over the course of my career, theoretical ideas have been tested in the crucible of the real world, and empirical experiences have given rise to more nuanced theoretical ideas. This was the case in my problem solving work, where a decade of teaching my problem solving courses served both as the "reality test" for my theoretical ideas and an inspiration for them. In studying teacher decision making, it is one thing to hypothesize the factors that shape teachers' choices; it is quite something else to try to model teachers' classroom behavior. In the work described in this article, the constant testing of our ideas against real data (videotapes of classroom teaching), combined with the need for theoretical clarity, is what produced ongoing refinements of our analytic scheme.

7.3 Observation 3: the need for multiple lenses, methods, and perspectives

This observation might be seen as a corollary to the first two, but it is worth highlighting on its own. Two illustrative examples are the bodies of research into teacher knowledge and teacher beliefs. From my perspective, inventories of teacher knowledge and/or teacher beliefs, on their own, can quickly become sterile. The question is not "what does a teacher know" or "what does a teacher say he or she believes" but, "how do a teacher's knowledge and beliefs play out in the classroom?" There is, for example, a corpus of research using questionnaires that examines teachers' beliefs (see Schraw & Olafson, 2002). However, what teachers say they believe and what they actually do in the classroom can be very different things (see, e.g., Cohen, 1990), and data from the questionnaires alone can be contradictory (as in Schraw & Olafson, 2002). Hence some form of triangulation is essential. An example of such triangulation is given in Swan (2006), where data from teacher questionnaires, student questionnaires, and independent observers' classroom observations are all juxtaposed and shown to be consistent. When such triangulation is done, one can have much greater confidence in the results.

I am convinced that research that lives in the tension between the theoretical and the empirical, and that employs multiple tools and perspectives, will ultimately enhance both theory and practice. I hope the example given in this paper indicates the ways in which theoretical and pragmatic lenses can be trained profitably on the same set of phenomena.

Acknowledgments This work was supported by the National Science Foundation (The Algebra Teaching Study, Grant DRL-0909815), to Alan Schoenfeld and Robert Floden, and the Bill and Melinda Gates Foundation (The Mathematics Assessment Project, Grant OPP53342). The work here truly represents a community effort, with significant contributions from Evra Baldinger, Danielle Champney, Aldo Dalla Piazza, Vinci Daro, Fadi El Chidiac, Christian Fischer, Denny Gillingham, Duanghathai Katwibun, Hee-jeong Kim, Mariana Levin, Nicole Louie, Sarah Nix, Dan Reinholz, Kim Seashore, Niral Shah, and Likun Sun from the University of California at Berkeley, and Rachel Ayieko, Adrienne Hu, Jerilynn Lepak, and Jamie Wernet from Michigan State University.

References

Algebra Teaching Study (2013, in preparation). *Beyond the numbers: What really counts in classroom evaluation schemes?* Berkeley, CA, and East Lansing, MI: University of California and Michigan State University, Algebra Teaching Study.

American Association of University Women. (1992). *How schools shortchange girls*. Washington, DC: AAUW and NEA.

Beeby, T., Burkhardt, H., & Caddy, R. (1980). *SCAN: Systematic classroom analysis notation for mathematics lessons*. Nottingham: England Shell Centre for Mathematics Education.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Cohen, D. (1990). A revolution in one classroom: The case of Mrs Oublier. *Education Evaluation and Policy Analysis, 12*(3), 311–329.

Cohen, E. G. (1994). *Designing groupwork: Strategies for heterogeneous classrooms* (Revised ed.). New York: Teachers College Press.

Common Core State Standards Initiative (2010). *Common core state standards for mathematics*. Downloaded June 4, 2010 from http://www.corestandards.org/the-standards.

PACT Consortium (2012) Performance Assessment for California Teachers. (2012) *A brief overview of the PACT assessment system*. Downloaded April 1, 2012, from http://www.pacttpa.org/_main/hub.php?pageName=Home.

Danielson, C. (2011) *The framework for teaching evaluation instrument*, 2011 Edition. Downloaded April 1, 2012, from http://www.danielsongroup.org/article.aspx?page=FfTEvaluationInstrument.

de Lange, J. (1999). *A framework for classroom assessment in mathematics* (Unpublished manuscript). Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science, Assessment Study Group.

Engle, R. A. (2011). The productive disciplinary engagement framework: Origins, key concepts and developments. In D. Y. Dai (Ed.), *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning* (pp. 161–200). London: Taylor & Francis.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399–483.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Understanding teaching and classroom practice in mathematics. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (Vol. 1, pp. 225–256). Charlotte, NC: Information Age Publishing.

Franke, M., & Webb, N. (2010). *Supporting student explanations: Variability in teacher practice across classrooms*. Manuscript in preparation. Available from authors, University of California, Los Angeles.

Franke, M., Webb, N., Chan, A., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education, 60*, 380–392.

González, N., Andrade, R., Civil, M., & Moll, L. (2001). Bridging funds of distributed knowledge: Creating zones of practices in mathematics. *Journal of Education for Students Placed at Risk (JESPAR), 6*(1–2), 115–132.

Henningsen, M., & Stein, M.K. (1997). Mathematical tasks and student cognition:Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, 28(5), 524–549.

Institute for research on learning (2011). *Accountable talk*. Downloaded November 26, 2011 from http://ifl.lrdc.pitt.edu/ifl/index.php/resources/principles_of_learning/.

Institute for Research on Policy Education and Practice. (2011). *PLATO (Protocol for language arts teaching observations)*. Stanford, CA: Institute for Research on Policy Education and Practice.

Junker, B., Matsumura, L. C., Crosson, A., Wolf, M. K., Levison, A., Weisberg, Y., & Resnick, L. (2004). *Overview of the Instructional Quality Assessment*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Marder, M., & Walkington, C. (2012) *UTeach Teacher Observation Protocol*. Downloaded April 1, 2012, from https://wikis.utexas.edu/pages/viewpageattachments.action?pageId=6884866&sortBy=date&highlight=UTOP_Physics_2009.doc.&.

Mathematics Assessment Project (2012). *Resources for assessment*. Downloaded January 10, 2012, from (http://map.mathshell.org/materials/index.php).

Measures of Effective Teaching (2012). Retrieved on July 4, 2012, from the Bill and Melinda Gates Foundation website: http://www.gatesfoundation.org/united-states/Pages/measures-of-effective-teaching-fact-sheet.aspx.

Moses, R. P. (2001). *Radical equations: Math literacy and civil rights*. Boston MA: Beacon Press.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics (NCTM). (1995). *Assessment standards for teaching mathematics*. Reston, VA: NCTM.

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.

National Research Council. (2005). How students learn: Mathematics in the classroom. Committee on how people learn, a targeted report for teachers. In M. S. Donovan & J. D. Bransford (Eds.), *Division of behavioral and social sciences and education*. Washington, DC: The National Academies Press.

Oakes, J., Joseph, R., & Muir, K. (2001). Access and achievement in mathematics and science. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (pp. 69–90). San Francisco: Jossey-Bass.

Pianta, R., La Paro, K., & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore: Paul H. Brookes.

Resnick, L., O'Connor, C., & Michaels, S. (2007). *Classroom discourse, mathematical rigor, and student reasoning: An accountable talk literature review*. Downloaded January 10, 2010, from http://www.learnlab.org/research/wiki/index.php/Image:Accountable_Talk_Lit_Review.pdf.

Schoenfeld, A. (1985). *Mathematical problem solving*. New York: Academic Press.

Schoenfeld, A. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31*(1), 13–25.

Schoenfeld, A. H. (2010). *How we think: A theory of goal-oriented decision making and its educational applications* (p. 2010). New York: Routledge.

Schoenfeld, A. H. (in press). American schools, today and tomorrow. In: O. L. Davis, Jr. & M. Kysilka (Eds.), *Schools for tomorrow's America*. Indianapolis, IN: Kappa Delta Pi.

Schraw, G., & Olafson, L. (2002). Teachers' epistemological world views and educational practices. *Issues in Education, 8*(2), 99–148.

Secada, W. (1992). Race, ethnicity, social class language, and achievement in mathematics. In: D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning,* pp. 623–660.

Shafer, M.C. & Romberg, T.A. (1999). Assessment in classrooms that promote understanding. In: E. Fennema & T.A. Romberg (Eds), *Mathematics classrooms that promote understanding* (pp. 159–184). Mahwah, NJ: Lawrence Erlbaum Associates.

Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning*. (CSE Technical Report 517). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340.

Stein, M. K., Grover, B., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488.

Swan, M. (2006). *Collaborative learning in mathematics: A challenge to our beliefs and practices.* London, England: National Research and Development Centre for Adult Literacy and Numeracy, and Leicester, England: National Institute of Adult Continuing Education.

University of Michigan (2006). Learning mathematics for teaching. *A coding rubric for measuring the mathematical quality of instruction* (Technical Report LMT1.06). Ann Arbor, MI: University of Michigan, School of Education.

Webb, N., and Romberg, T.A. (1992). Implications of the NCTM Standards for Mathematics Assessment. In T.A. Romberg (Ed.), *Mathematics Assessment and Evaluation: Imperatives for Mathematics Education*, (pp.37–60). Albany: State University of New York Press.

Zevenbergen, R. (2000). "Cracking the code" of mathematics classrooms: School success as a function of linguistic, social, and cultural background. In J. Boaler (Ed.), *Multiple perspectives on mathematical teaching and learning* (pp. 201–224). Westport, CT: Ablex Publishing.