Title
Making sense of teaching

Permalink
https://escholarship.org/uc/item/7rh5k7nt

Journal
ZDM – Mathematics Education, 48(1-2)

ISSN
1863-9690

Author
Schoenfeld, Alan H

Publication Date
2016-04-01

DOI
10.1007/s11858-016-0762-3

Peer reviewed

CrossMark

COMMENTARY PAPER

Making sense of teaching

Alan H. Schoenfeld[1]

Abstract This article begins with a series of meta-level comments on the enterprise of observing and theorizing classroom teaching, using as a springboard a frequently discussed model from the literature. It then provides a series of commentaries on the empirical articles in this issue (volume 48, issue 1) of ZDM, with a main focus on methodological issues. It concludes with a brief return to the meta-level, with a comment on the field's use of the term "model".

Keywords Teacher proficiency · Classroom observations · Theory · Research methods · Models

1 Introduction

Some thirty years ago Lee Shulman visited Berkeley to talk about contemporary research on teaching. He discussed the "process–product paradigm", in which researchers counted frequency and duration of particular activities engaged in by teachers, and sought correlations with student outcomes. Shulman bemoaned the state of the art, saying that the field needed deeper explanations of teacher proficiency and better methods for studying teaching. Soon afterward, he introduced the concept of teachers' pedagogical content knowledge (Shulman, 1986), in an attempt to focus on the underpinnings of teacher proficiency. This work opened up new lines of inquiry that flourish to this day, as reflected

✉ Alan H. Schoenfeld
    alans@berkeley.edu

[1] Elizabeth and Edward Conner Professor of Education, Graduate School of Education, University of California, Tolman Hall, MC 1670, Berkeley, CA 94720-1670, USA

in this volume. Shulman was not alone: Winne (1987), for example, wrote:

> Research on teaching primarily adopts the process–product paradigm. Within this paradigm, researchers often speculate about cognitive operations that students engage in during instruction as a means to explain *how* teacher behaviors (processes) correlate with or cause student achievement (products). This paper argues that the methodology of process–product research is (1) ill-suited to generating theories of teaching effectiveness that use students' cognition to explain process–product relations and (2) invalid for testing such explanations. (Winne, 1987, p. 333)

What a difference three decades can make! Reading the current volume of *ZDM* provides significant evidence of how far the field has come over the past three decades. While there is always continuity in research and the field builds on what has come before, the space of theorizing, research methods, and data gathering has expanded tremendously—and with that expansion the ability to provide increasingly deep explanations of the cultural contexts in which teaching takes place, the character of teacher knowledge and decision making, the properties of rich learning environments, and student learning have all grown in important ways. Indeed, it is hard to imagine a volume like the current one being produced even a half dozen years ago. Issues are far from settled, but the field is in a productive state of ferment.

In what follows I begin with some meta-level comments on the enterprise of observing and theorizing classroom teaching, using as a springboard a model discussed in the theoretical chapter by Schlesinger and Jentsch. I then discuss the empirical chapters from this volume, in order, elaborating as I do on some of the methodological themes

Fig. 1 Three-dimensional model of instructional quality (reproduced from Klieme & Rakoczy, 2008, p. 228)
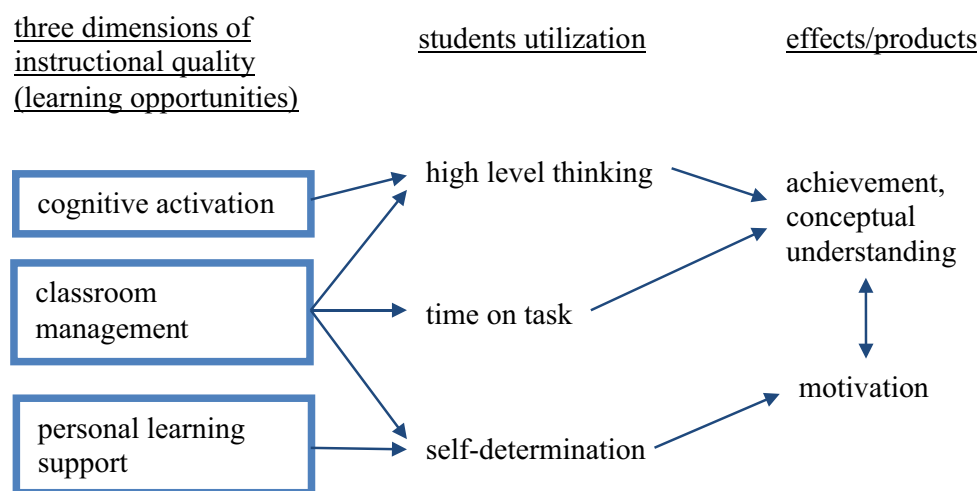
1.1 Some meta-level comments on the enterprise

In "Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations", Schlesinger and Jentsch address the question of measuring instructional quality, which raises deep questions of theory as well as method. I will take their jumping-off point, a three-dimensional model of instructional quality drawn from Klieme and Rakoczy (2008), as my own jumping-off point for problematizing the very question of what we mean by "instructional quality" and, more broadly, what we mean by the outcomes of instruction. As seen in Fig. 1, the authors posit three "independent variables" describing the affordances of instruction (learning opportunities), cognitive activation, classroom management, and personal learning support. These are mediated by the students' utilization of the affordances of the environment (the middle column of Fig. 1) and produce the dependent variables or effects/products: achievement, conceptual understanding, and motivation. It is clear that each of the independent variables is important and can be measured in some way, as can the effects or products. Classroom management, for example, is clearly a necessary but not sufficient condition for effective learning: a classroom in chaos is not going to support powerful learning, but a well-organized classroom that focuses on low-level mathematics will not do so either. In short, this is a perfectly reasonable kind of model. My point in what follows is not to challenge the model, but to problematize the very enterprise in theoretical terms.

I begin with the effects and products. What are the outcomes of instruction, and how should they be measured?

To illustrate the issue, I begin with a question I have been asked by many Deans at colleges and universities in the USA. Faced with budgets that include a large number of recitation sections taught by faculty and with 30 students in each, the administrators wonder if lecture classes would serve as well. They pose the simple question, "Is there a meaningful difference between instruction in large lecture classes and small classes taught by faculty?"

My response is, "It depends on what you value". Suppose, for example (as has been the case at many universities in the USA), that the final examination in the course is a multiple choice test that focuses largely on procedural skills.[1] If one of the most effective teachers in the department is put in charge of instruction, and prepares the teaching assistants well, the students will do as well, on average, on a straightforward multiple choice exam as students taught by faculty in smaller sections. If the content is more challenging—say it includes modeling and problem solving—and the final exam includes "essay questions" in which the students are asked to demonstrate their understandings, then students in the smaller classes may do better. But, there are other outcomes as well. In universities across the USA, as many as 5 % of entering students think they may become mathematics majors, but this percentage dwindles substantially as students progress through calculus. In contrast, in a small number of colleges where significant attention is given to calculus instruction in small sections, there is a substantial *increase* in the percentage of students intending to major in mathematics (Schoenfeld, 2000). Or, consider the fact that in the U.S., a significant proportion of the population says that they hate

---

[1] It goes without saying that good multiple choice tests can be created. But the ones I have seen in practice, created by faculty without special training in test construction, have not been very sophisticated.

mathematics (for evidence, google "hate math" or "what % of the population hates math?"). Such dislike is, of course, one of the outcomes of people's experience with mathematics in school. Does this affective outcome (which is very consequential in societal terms) belong in a model of instructional quality? If so, how does one measure for it?

So the first question is, what does one value as outcomes? Once that issue has been decided, then there is the question of how such outcomes are measured. "Achievement" and "Conceptual understanding" as identified in Fig. 1 may seem to be straightforward and well defined, but they are not. The fact that the teams writing the TIMSS and PISA exams spent so much achieving consensus on the content of the exams—and, that the national rankings on the two exams were similar but far from identical—indicates that each exam was an indication of the mathematical values of its creators.

In the USA we have had dramatic examples during the "math wars", when some instruction (and exams) focused primarily on skills while others examined skills, concepts, and problem solving. For example, the SAT-9 was a state-wide examination in California that, for many years, was used as California's official outcome measure of student learning. The MARS tests, created by the Balanced Assessment Project, were designed to test skills, concepts, and problem solving. Ridgway, Crust, Burkhardt, Wilcox, Fisher, and Foster (2000) compared more than 16,000 students' performance at Grades 3, 5, and 7 on the two examinations. For each student, Ridgway et al. reported the scores on each exam as being "proficient" or "not proficient". The results are given in Table 1.

Any two mathematics tests are likely to correlate at some level, and that is clearly the case here. But, in all three grades a large number of students who were declared proficient on the official statewide test of skills were declared "not proficient" on a test that included assessments of

conceptual understanding and problem solving. How one chooses to assess the desired outcomes really matters!

Let us now turn to the input side. As noted above, each of the learning opportunities in Fig. 1 is clearly important. Yet, there are many questions one can ask. Are the categories identified comprehensive? That is, is anything important missing? Can each of the items be operationalized in reasonable ways? Can we recognize them, and measure them, in ways that are well defined and that capture what counts? Are they multidimensional constructs, so that a model that looks simple in description is in fact extremely complex when one tries to operationalize it? One could imagine very different and very complex characterizations, for example, of cognitive activation. At a meta-level, a major goal of construct definition and measurement is to have terms that are well defined and measurement tools that capture those definitions in meaningful ways.

The same is true for the middle dimension, student utilization. Note that "time on task" is a simple measurable variable, while "high level thinking" is a very complex concept—one, fortunately, that we have, as an intellectual community, made great progress on over the past decades. (But, it too, is multi-dimensional). The term "self-determination" is an amalgam, which leads me to wonder how complex it will be when it is operationalized.

Finally, one wonders about the purposes to which any model will be put. Are the intentions descriptive, in which the case complexity of individual terms may not be that important an issue (especially if the purpose is to frame research in the large)? Or, is the framework intended to support action aimed at improvement, in which case complexity and/or lack of specificity could hamper productive use of the framework?

My point here has not been to critique the framework in Fig. 1, but rather to use it to illustrate the kinds of challenges we face as a field if we hope to make productive use of analytic frameworks and their associated tools for research. My colleagues and I have been working on a set of criteria for classroom observation frameworks (Schoenfeld, Floden, and the Algebra Teaching Study and Mathematics Assessment Projects, in preparation). We recognize that different frameworks for conceptualizing classroom activities will have different purposes, e.g., for research, for professional development, and/or teacher evaluation. Even so, we argue that there are some properties that all frameworks for conceptualizing classroom activities should have. Those properties are summarized in Table 2.

My research groups' theoretical and observational framework, the Teaching for Robust Understanding (TRU) framework, is discussed at length in a series of papers (Schoenfeld, 2013, 2014, 2015). TRU looks very different from the kind of structure in Fig. 1. First, it involves a shift

Table 1 Comparison of students' performance on two examinations

MARS	SAT-9	
	Not proficient	Proficient
Grade 3 ($N = 6136$)		
Not proficient (%)	27	21
Proficient (%)	6	46
Grade 5 ($N = 5247$)		
Not proficient (%)	28	18
Proficient (%)	5	49
Grade 7 ($N = 5037$)		
Not proficient (%)	32	28
Proficient (%)	2	38

Reproduced with permission from Ridgway, Crust, Burkhardt, Wilcox, Fisher, & Foster, (2000)

Table 2 Desirable properties of classroom observation frameworks

A framework for conceptualizing classroom activities should be
Grounded in a theory of domain proficiency, which characterizes "necessary and sufficient" conditions for proficient domain performance
Parsimonious, in that the framework helps focus on what truly counts—preferably in ways that are easy to understand and remember
Fruitful (and timely) for the intended purposes
Quantitatively robust, both in measurement terms and in empirical studies of how it correlates to powerful student performance

Reproduced, with permission, from Schoenfeld, Floden, et al., in preparation

of frame, in that its primary orientation is to how the student experiences the classroom. Second, it posits that five dimensions of classroom activities are necessary and sufficient to characterize and measure what counts in classrooms. The framework and the tools are available at http://ats.berkeley.edu and http://map.mathshell.org. My purpose here is not to elaborate on the TRU framework, but to note that there are very different ways to examine the same territory. Conversations about the meta-level issues discussed here will, I hope, be good for the field.

1.2 Notes on the empirical papers, with a focus on method

With those meta-level issues as backdrop, let me turn to the empirical studies in the volume, in the order they appear. In "Diagnostic competence of primary school mathematics teachers during classroom situations", Jessica Hoth, Martina Döhrmann, Gabriele Kaiser, Andreas Busse, Johannes König, and Sigrid Blömeke describe a video follow-up to TEDS, called TEDS-FU, which uses video stimuli to present teachers with various pedagogical scenarios, and explore the teachers' reactions. Here I must stop, in the spirit of my opening comments, to note the novel character of the content, the methods, and the findings. Although "diagnostic teaching" has its roots going back to the 1980s (Bell, 1993; Bell et al., 1988; Weinert, 1990), the concept languished for many years until it was resuscitated as "formative assessment" (see, e.g., Black & Wiliam, 1998). The key idea, that what matters for learning is far more than whether the mathematics is correct—what matters is what sense the student makes of the mathematics—is well known. There is a good argument that effective teaching involves "meeting the students where they are" and helping students both to build on solid foundations and to reflect on misconceptions. But the spread of that perspective has not been rapid. Hence, a focus on teachers' capacity in this area is timely. The means of exploring teachers' understandings in this article is quite typical of this volume, which is what I find so nicely revolutionary: if you want to understand teachers' thinking about classroom situations, present them with videos of such situations and see what they do! This may seem natural to readers at this point, but it represents a

major change in the direction of ecological validity. That is a fundamentally positive development. Finally, the findings reflect a current and important reality. For teaching to be fully effective, the teacher must be responsible to the mathematics *and* responsive to current student understandings. Understanding what teachers see, and how to help them achieve that responsibility and responsiveness, in a significant challenge.

"Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic", by Lena Pankow, Gabriele Kaiser, Andreas Busse, Johannes König, Sigrid Blömeke, Jessica Hoth, and Martina Döhrmann, documents the fact that as situations become more complex, those who give correct analyses or answers tend to take more time, while those who give incorrect answers tend to respond more quickly. This is important in terms of training and mindset. In the words of George Pólya, "First. You have to understand the problem". (Pólya, 1945, overleaf). I remember reading, many years ago, a study that examined how long calculus students, and their instructors, took to read problems before they started solving them. Interestingly, the instructors—who surely knew the material—took three times as long, on average, before they started working on a solution.[2] That is because they understood that they needed to have a solid grasp of what the problem asked for before they started working on its solution.

"Instructional reasoning about interpretations of student thinking that supports responsive teaching in secondary mathematics", by Elizabeth Dyer and Miriam Gamoran Sherin, represents another step toward ecological validity, in two important ways. First, the video segments under discussion are instances of teachers' own practice, which are clearly meaningful in ways that paper-and-pencil scenarios, or even videos of other teachers' practices, are not. Second and perhaps more important, the fact that the teachers made the choices provides significant insight into what the teachers themselves saw as interesting and important. This is a direct window into teacher thinking, as opposed to

---

[2] Unfortunately, my filing system is not as good as it should be. I could not find the specific reference.

questionnaires about teacher beliefs or even post hoc analyses of classroom videos. To paraphrase *How People Learn* (Bransford, Brown, & Cocking, 2000), p. 26: Learners are learners, and what applies to children's learning applies to adult learners as well. In particular, responsive teaching requires having a sense of what the learner understands—even when the learner is a teacher. The tools and techniques of this article give us closer access to those understandings.

The next article, "Uncovering predictors of disagreement: Ensuring the quality of expert ratings" by Jessica Hoth, Gabriele Kaiser, Andreas Busse, Martina Döhrmann, Johannes König, and Sigrid Blömeke, moves into different territory, focusing on methodological issues; it deals with differences in expert ratings of various video-based test items from the TEDS-FU study. This article serves as a reminder that ratings are not a matter of objective fact, but a matter of values; moreover, sometimes a community has consistent values, and sometimes it does not. Many years ago, I worked with a person named Renee who had some fundamental difficulties with the concept of slope. She had been taught the topic numerous times, but the basic ideas never seemed to take hold—she had certain difficulties she could not overcome. Renee came into my lab, and worked with me for 2 hours in a very exploratory session, using a new graphing tool my research group had developed. [The session is described in Malcolm Gladwell's (2008) *Outliers*, pp. 239–247]. At the end of the 2 h she had an epiphany—she finally figured out how things fit together. I was ecstatic, because the episode represented a true sense making experience—one that would stay with her, unlike her previous instruction.

Not long after that experience, a colleague of mine came to visit. She was a learning expert; some years before we had been postdocs together. I showed her the videotape of my session with Renee. Her reaction was this: "You know, Alan, when we design our instruction we work with master teachers. We learn a lot from what they do". That is, she thought that I had wasted a lot of time with Renee; if I had simply shown her the mathematics in a more straightforward way, she would have learned more efficiently. From my colleague's perspective, what I saw as a fundamental (and necessary, given her history) act of sense making on Renee's part was, instead, inefficient teaching on my part. Again: values matter. Sometimes an item can be improved or clarified, in which case experts may agree; but sometimes, a difference in value systems, even between experts, may make agreement impossible. Or, as indicated in this article, sometimes different communities have different values. University faculty and teachers may simply tend to see things in different ways.

"Further exploration of the classroom video analysis (CVA) instrument as a measure of usable knowledge for teaching mathematics: taking a knowledge system perspective", by Nicole Kersting, Taliesin Sutton, Crystal Kalinec-Craig, Kathleen Jablon Stoehr, Saeideh Heshmati, Guadalupe Lozano, and James Stigler, is yet another example of the use of video clips as an ecologically meaningful prompt, and the use of a variety of quantitative tools to validate the measures and explore their utility. Likewise, "Measuring mathematical teachers' professional competence by using video clips (COACTIV video)", by Georg Bruckmaier, Stefan Strauss, Werner Blum and Dominik Leiss, explicitly addresses teachers "situated reaction competency" by examining teachers' reactions to stimulus videotapes. The authors note explicitly that such efforts are in stark contrast to the most common approach to measuring teacher competencies, using paper-and-pencil tests. The authors also discover a notable difference in teaching approaches between Gymnasium and Hauptschule teachers, suggesting that the latter "refrained from student-oriented answers, because they often experience students in their teaching that are not capable of working independently". I would like to suggest that the issues may be more complex.

In the research reported in Schoenfeld (1988), I spent a year in the classroom of a teacher who laid out the content of instruction very carefully for students, in a step-by-step manner. One day I asked him if he had ever thought of giving his students a problem and letting them play with it. "No", he said, "that would just confuse them. I do that with my honors students". I visited his honors class, and that is what he did. What this indicates is that his beliefs were context-specific. He possessed the relevant pedagogical content knowledge, but he only used it in contexts where he felt it was appropriate.

I find it interesting that in their article "Epistemological beliefs of prospective preschool teachers and their relation to knowledge, perception, and planning abilities in the field of mathematics: a process model", Simone Dunekacke, Lars Jenßen, Katja Eilerts, and Sigrid Blömeke venture into preschool. This is a useful extension of the literature, because we know much less about preschool mathematics teachers' mathematical/pedagogical understandings than we do about teachers of K-12. It is reasonable to use questionnaires to gain baseline information about teacher practices. One can hope that in a few years there will be enough information to allow for the kinds of more ecologically valid research discussed in some of the other chapters. I do note that the authors claim that there has not bee a consistent definition of beliefs – but, there has been progress in that direction: see, e.g., Li and Moschkovich (2013).

With "Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME)" by Johannes König and Charlotte Kramer, we return once again to the theme of ecological validity, and to the issue

of values. The authors, like others, make the point that videotapes of classroom excerpts provide a much richer and more ecologically valid characterization of classroom episodes than is possible with paper and pencil measures. The authors suggest that "classroom management expertise, compared with general pedagogical knowledge, is much more dependent on the expertise level acquired during professional development, whereas general pedagogical knowledge can be acquired as early as during the theoretical initial teacher education at university". To this I would add two caveats. The first is that their statement is likely to be accurate at present, given the current forms of teacher preparation and professional development. I have seen beginning teachers who were remarkably good at classroom management, because they came from a teacher preparation program that focused so intently on individual student learning that the teachers were remarkably attuned to their students' understandings, and as a result had very few challenges with classroom management.

The second is that what constitutes good classroom management is itself a matter of values. My research group has found itself judging classrooms differently than another classroom observation framework, which tends to favor orderly classrooms of the "demonstrate and practice" mode (Schoenfeld, Floden, et al., in preparation). The other observation rubric rated a classroom low on classroom management, because the classroom was noisy for some time and the teacher did not provide explicit guidance about what the students should be doing. In contrast, the TRU rubric rated it as scoring well, because the noise was produced by students engaging in animated mathematical sense making. The discussions were not yet coherent, because this was the early part of the lesson. The students were—as intended—struggling with the content, which made for conceptual messiness and some noise. But, the students' grappling with ideas led to fertile conversations later in the lesson. From our perspective, the noisy engagement was a precursor of progress; from the other rubric's perspective, it represented a lack of orderly classroom management. Thus, even the question of what one considers out to be "effective classroom management" is a matter of values. Indeed, in our experience, one of the great challenges of professional development has been to support teachers in living with the "messy" part of our Formative Assessment Lessons (see http://map.mathshell.org/lessons.php) and deter them from trying to "fix" student confusions by telling them the right answers.

"The role of perception, interpretation, and decision making in the development of beginning teachers' competence" by Rossella Santagata and Cathery Yeh has many of the virtues that I have highlighted about this issue of *ZDM* in general. In particular, issues of both ecological validity and of triangulation are nicely handled in this article,

with its use of three longitudinal case studies that include a classroom video analysis survey, classroom observations and interviews about teachers' instructional decisions, and whole-day shadowing. Multiple data sources allow for minimizing the likelihood that some of the findings are artifactual, and the connections to the literature (e.g., Blömeke et al.'s (2015) conceptualization of teacher competence) and refinements of it add to the cumulative, connected nature both of this volume and of the progress of the field as a whole.

In "Using multimedia questionnaires to study influences on the decisions mathematics teachers make in instructional situations", Patricio Herbst, Daniel Chazan, Karl W. Kosko, Justin Dimmel, and Ander Erickson continue their innovative explorations into the factors that shape teachers' decision making. The paper expands their description of things that matter, including teachers' sense of instructional norms and their perceived obligations to the discipline. I shall again focus on research methods, however, because their research paradigm is the exception that probes the rule with regard to the issue of ecological validity. An interesting question is, when is "real" too real? In general, a video of a classroom situation serves as a much better stem for the question "what would you do next" than a body of text describing the classroom situation; that is the reason so many video prompts are used in the studies reported in this volume. But, videos also contain a huge amount of information, some of which may be distracting or not relevant. The issues explored by Herbst, Chazan, and colleagues have to do with the existence of norms. A real classroom video necessarily focuses attention on the behavior of the particular teacher in the video, which may be a distraction. The use of a less detailed scenario that still captures the gist of the issue (in this case, whether a particular teaching decision violates pedagogical obligations to the discipline) allows viewers to focus on the issue without being distracted by what would, in this case, be irrelevant detail. The animated scenario is a realistic description of the situation being explored, although not "real". That is, the animation is valid for the intended research purposes.

"Responding to children's mathematical thinking in the moment: an emerging framework of teaching moves", by Victoria Jacobs and Susan Empson, follows an established tradition of unpacking aspects of expert practice in order to make that practice more accessible to those who wish to employ it. What the authors term "responsive teaching" is an aspect of teaching that is becoming increasingly important, and understanding the teacher moves that support it is a worthy enterprise; here too, multiple sources of data (interviews and class lessons) inform the work. Also of methodological interest, "Instructional decision making and agency of community college mathematics faculty" by Elaine Lande and Vilma Mesa makes use of systemic

functional linguistics to document linguistic usage on the part of community college faculty (who often hold part-time, untenured positions) when commenting on animations of the type described in the article by Herbst, Chazan, and colleagues. As discussed above, the prompts are more useful for not offering distracting detail.

2 Discussion

I want to begin my discussion by making a point about language use. Many of the authors in this volume used the term "model". One sees particular models in Fig. 1 in the article by Schlesinger and Jensch, Fig. 1 in the chapter by Hoth et al., Fig. 3 in the article by Santaga and Yeh, Fig. 1 in the article by Jacobs and Empson, and Fig. 1 in the chapter by Lande and Mesa. Each of these figures depicts certain relationships, mostly using objects and arrows. Although there is a long tradition of such usage, I find the way in which the term is used troubling, especially in mathematics education.

In mathematics, and in mathematics education—there is, after all, an entire subfield devoted to the didactics of mathematical modeling[3]—a mathematical model is a precise and well-defined representation of a system of objects and relationships. Objects in the model correspond to objects in the system being represented, and the (hypothetical) relationships between the objects are specified in detail as well. To give a few examples:

- A gravitational model of the solar system includes some representation of the planets that captures their locations, mass, and the directions and speed in which they are traveling at some particular time; it also includes a representation of the laws of gravitational attraction, which enable one to compute the projected locations of the planes after some time has elapsed.
- A mathematical model of temperature flow in a laminar plate uses differential equations (the representation of relationships) to compute the temperature at different points in the plate, given particular initial conditions.
- The models of teacher decision making in Schoenfeld (2011) provided explicit explanations of teachers' decision making, on the basis of their resources, orientations, and goals.

I would be much happier if the field restricted its use of the term "model" to mathematical models of this type,

and used terms such as "representation" or "description" for schematic illustrations such as the ones in this volume. But, that is a small complaint. I want to conclude by noting, once again, the tremendous amount of progress the field has made in conceptualizing and investigating teacher decision making. This issue of *ZDM* stands as testimony to the advances that have been made, not only since the observations Lee Shulman made some 30 years ago, but in the past few years, as research has become more methodologically sophisticated and more focused on "what counts".

References

Bell, A. (1993). Some experiments in diagnostic teaching. *Educational Studies in Mathematics, 24*(1), 11–137.

Bell, A., Swan, M., Onslow, B., Pratt, K., & Purdy, D. (1985). *Diagnostic teaching: teaching for long term learning*. Nottingham: Shell Centre for mathematical education.

Bishop, A. J. (1976). Decision-making, the intervening variable. *Educational Studies in Mathematics, 7*, 41–47.

Bishop, A. J., & Whitfield, R. C. (1972). *Situations in teaching*. London: McGraw-Hill.

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

Blömeke, S., Gustafsson, J. E., & Shavelson, R. (2015). Beyond dichotomies: viewing competence as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (2000). *How people learn* (Expanded ed.). Washington, DC: National Academy Press.

Bruckmaier, G., Krauss, S., Blum, W., & Leiss, D. (2016). Measuring mathematical teachers' professional competence by using video clips (COACTIV video). *ZDM Mathematics Education, 48*(1) (this issue).

Dunekacke, S., Jenßen, L., Eilerts, K., & Blömeke, S. (2016). Epistemological beliefs of prospective preschool teachers and their relation to knowledge, perception, and planning abilities in the field of mathematics: a process model. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0711-6 (this issue).

Dyer, E.B., & Sherin, M.G. (2016). Instructional reasoning about interpretations of student thinking that supports responsive teaching in secondary mathematics. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0740-1 (this issue).

Gladwell, M. (2008). *Outliers: the story of success*. New York: Little, Brown.

Herbst, P., Chazan, D., Kosko, K.W., Dimmel, J., & Erickson, A. (2016). Using multimedia questionnaires to study influences on the decisions mathematics teachers make in instructional situations. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0727-y (this issue).

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM Mathematics Education, 48*(1) (this issue).

Hoth, J., Schwarz, B., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Uncovering predictors of disagreement: ensuring the quality of expert ratings. *ZDM Mathematics Education, 48*(1) (this issue).

Jacobs, V.R., & Empson, S.B. (2016). Responding to children's mathematical thinking in the moment: an emerging framework of teaching moves. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0717-0 (this issue).

---

[3] See http://www.icmihistory.unito.it/ictma.php for a history of ICTMA, The International Community of Teachers of Mathematical Modelling and Applications.

Kersting, N.B., Sutton, T., Kalinec-Craig, C., Jablon Stoehr, K., Heshmati, S., Lozano, G., & Stigler, J.W. (2016). Further exploration of the classroom video analysis (CVA) instrument as a measure of usable knowledge for teaching mathematics: taking a knowledge system perspective. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0733-0 (this issue).

König, J., & Kramer, C. (2016). Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0705-4 (this issue).

Lande, E., & Mesa, V. (2016). Instructional decision making and agency of community college mathematics faculty. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0736-x (this issue).

Li, Y., & Moschkovich, J. (2013). *Proficiency and beliefs in learning and teaching mathematics*. Rotterdam: Sense Publishers.

Pankow, L., Kaiser, G., Busse, A., König, J., Hoth, J., Döhrmann, M., & Blömeke, S. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM Mathematics Education, 48*(1) (this issue).

Ridgway, J., Crust, R., Burkhardt, H., Wilcox, S., Fisher, L., & Foster, D. (2000). *MARS report on the 2000 tests*. San Jose: Mathematics Assessment Collaborative.

Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM Mathematics Education, 48*(1). doi:10.1007/s11858-015-0737-9 (this issue).

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM Mathematics Education, 48*(1) (this issue).

Schoenfeld, A. H. (2000). Purposes and methods of research in mathematics education. *Notices of the American Mathematical Society, 47*(6), 2–10.

Schoenfeld, A. H. (2011). *How we think: a theory of goal-oriented decision making and its educational applications*. New York: Routledge.

Schoenfeld, A.H. (2013). Classroom observations in theory and practice. *ZDM—The International Journal of Mathematics Education, 45*, 607–621. doi:10.1007/s11858-012-0483-1.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? *Educational Researcher, 43*(8), 404–412. doi:10.3102/0013189X1455.

Schoenfeld, A. H. (2015). Thoughts on scale. *ZDM Mathematics Education, 47*, 161–169. doi:10.1007/s11858-014-0662-3.

Schoenfeld, A.H., Floden, R.E., and The Algebra Teaching Study and Mathematics Assessment Projects. (2016). On classroom observations. Manuscript available from the first author (in preparation).

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

TeachingWorks project. (2016). High leverage practices. http://www.teachingworks.org/work-of-teaching/high-leverage-practices.

Weinert, F. E., Schrader, F. W., & Helmke, A. (1990). Educational expertise. Closing the gap between educational research and classroom practice. *School Psychology International, 11*, 163–180.

Winne, P. (1987). Why process-product research cannot explain process-product findings and a proposed remedy: the cognitive mediational paradigm. *Teaching and Teacher Education, 3*(4), 333–356.