New Ways to Measure Adult Developmental Differences Among Teachers[1]

by James K. L. Hammerman, TERC, Cambridge, MA
Rebecca Mitchell, TERC & Harvard Graduate School of Education, Cambridge, MA

Introduction

Teacher professional development can be extremely effective in transforming teaching practice and improving schools (Ball & Rundquist, 1993; Brahier & Schaffner, 2004; Hawley & Valli, 1999; Kent, 2004). In general, effective professional development is "situated" in the complexities of classroom teaching (Ball & Cohen, 1999), focuses on subject matter, pedagogy and subject-specific pedagogy (Borko & Putnam, 1995), involves teachers in "defining and shaping the problems of practice" to be addressed (Lieberman, 1995), builds leadership capacity and a learning community among teachers and gives them a chance to learn content in the inquiry-oriented ways that they'll use with students (Mundry & Dunne, 2003). What's often not addressed in this literature is the different ways that teachers experience the same professional development program and how that affects program effectiveness. There's a strong need for theory which may help us understand the differential effectiveness of professional development programs.

One such theory describes the qualitatively different ways that adults construct their understanding of the world, including their ability to take and coordinate others' perspectives, their locus of authority, and the complexity of their thinking (Kegan, 1982; , 1994). Prior small scale case study research (Hammerman, 2002) suggests that these types of constructive-developmental differences among adults affect how and what teachers learn in professional development (PD) programs: Specifically, Hammerman found that some teachers want programs to tell them how to teach in new and improved ways but only "implement" PD ideas when they receive strong school-based support. Other teachers are more skeptical even if their basic philosophy tends to agree with that of the PD program. They *don't* want to be told what to do, but want to reflect on and work to find ways to integrate the ideas of the PD program into their own personally generated teaching practice that responds to the needs of their students independent of whether there is school-based support. Deeper knowledge of this theoretical perspective and its implications for teacher learning could lead to changes in the design, and ultimately the effectiveness, of teacher professional development programs.

To accomplish this, however, we need further empirical research on the impact of the match or mis-match between teachers' constructive-developmental level and what programs offer and expect from teachers. This research could have several components, including curriculum analyses of the constructive-developmental demands and supports

---

[1] Support for this paper was provided by the National Science Foundation (NSF) under grant ESI-0439281, *Exploratory Research to Develop Methods for Studying Adult Development in Math Teacher Professional Development* at TERC, Cambridge, MA. The views expressed herein are those of the authors and do not necessarily reflect those of the NSF.
Please cite as: Hammerman, J. K. L., & Mitchell, R. (2006). *New Ways to Measure Adult Developmental Differences Among Teachers*. Paper presented at the American Educational Research Association (AERA), San Francisco, CA. Available at: http://www.terc.edu/work/133.html.

of professional development programs, and broader-scale work on how constructive-developmental differences affect teachers' experiences of these programs. The latter research strand requires accurate and easy measurement of teachers' constructive-developmental levels. The accepted reliable measure of constructive-developmental difference in Kegan's theory is the Subject-Object Interview (SOI) (Lahey, Souvaine, Kegan, Goodman, & Felix, 1988). This intensive, open-ended interview measure requires extensive specialized training to conduct and score, and is extremely time-consuming (several hours to conduct, transcribe, and score each interview). Therefore, it is not well-suited for the type of medium- to large-scale research suggested here.

Through a grant from the National Science Foundation, the *Exploratory Research to Develop Methods for Studying Adult Development in Math Teacher Professional Development* project (ESI-0439281) is trying to develop alternative, more efficient measures of constructive-developmental difference that could be used in larger scale studies of teacher professional development. This paper describes our efforts to develop such instruments and some of the difficulties we've encountered.

Theoretical background[2]

Kegan's constructive-developmental theory (1982; 1994) describes five qualitatively different meaning-making structures or Orders of Mind that occur across the life span. Each centers around a different balance between those aspects of ourselves and our world that we can work with, relate to, and have some control over (aspects that Kegan calls "object"); and aspects of ourselves and the world that we are made up by, that provide the lens or frame through which we see, that we don't have perspective on because they constitute what we *are* (aspects that Kegan calls "subject"). It is the shifting in these "subject-object" balances that represents development. The larger the realm of what we consider "object," the more we can take responsibility for and the more complex is our understanding of ourselves in the world.

Young children in Kegan's First Order are subject to their perceptions—when something *looks* different, to them it *is* different. Older children develop a sense of "durable categories" in the physical and social worlds. At the Second Order, people can coordinate their impulses and perceptions to describe their own and others' enduring dispositions and preferences: "I'm a friendly person not just because I feel happy or have a friend today, but because I've noticed that's true about me in lots of situations and circumstances." They can coordinate changing perceptions over time to conserve concrete quantities like volume. They develop a point of view because they now know (where before they did not) that not everyone sees the world the same way. They care about how others perceive them because those perceptions may have concrete consequences for them.

Most people gain some perspective on these "durable categories" to construct an understanding that can coordinate and integrate them within larger cognitive and social principles—Kegan's *Socializing* Third Order—though many adults (roughly 1/8 to 1/3,

[2] This description of Kegan's theory is based on that in Hammerman, 2002.

see Figure 1) continue to make meaning in whole or in part in Second Order ways. In the Third Order, people can coordinate several points of view within a sense of their own role within a social structure. They can internalize others' perspectives and thus, care about others' opinions of them as such, not just for how those opinions shape others' actions towards them. People at this Order can use abstractions and inference to coordinate concrete data, and can develop hypotheses and respond to abstract ideals and values. The idea of doing things "because it's the right thing to do" even if it's not in your own self-interest makes sense at this Order of Mind. Kegan describes this meaning-making structure as "Traditionalist" or "Socializing" in that it includes an internalized sense of mutual reciprocity in social relationships, and therefore enables people at the Third Order to be responsible for their own role within a larger social structure. Most adults (between 5/9 and 2/3, see Figure 1) make meaning in whole or in part in this way and, given contexts which provide appropriate external models for belief and action, adults using this meaning-making system can be quite successful.[3]

However, Kegan (1994) argues that our society often demands something more from adults, and these demands serve to move people towards the Fourth Order. In this "Modernist" or "Self-authoring" Order of Mind, adults come to coordinate their multiple roles and the different expectations others hold for them within their own self-generated, relationship-regulating frameworks. Someone at the Third Order might be torn apart by

[3]The statistics I report here are compiled from Kegan's (1994) description of the distribution of Subject-Object Interview scores for 282 people drawn from 12 dissertation studies (pp. 188-197). Proportions of people at each Order are correlated, in part, with age, education, and socio-economic status; thus the rough ranges presented in the text. The full dissertation sample reported by Kegan is biased towards a professional and well-educated population, though he also describes a subset of three studies (N=75) that represents a more complete SES range.

In Figure 1, below, I chart distributions of Orders of Mind for both of these samples—it is not clear which better describes the population of teachers. Those at transitional Orders (2-3, 3-4, 4-5) make meaning in ways that are sometimes like the Order below, and sometimes like the Order above—a mix of both Orders. I have tried to represent this by visual patterns that overlap the two related Orders.

To determine what percentage of people make meaning, at least in part at a particular Order, add the percentages at the transitions below and above to the percentage at the Order itself. Thus, for example, 54% of the Dissertation Composite (8% in the 2-3 Transition + 14% at the Third Order + 32% in the 3-4 Transition) make meaning at least in part at the Third Order.

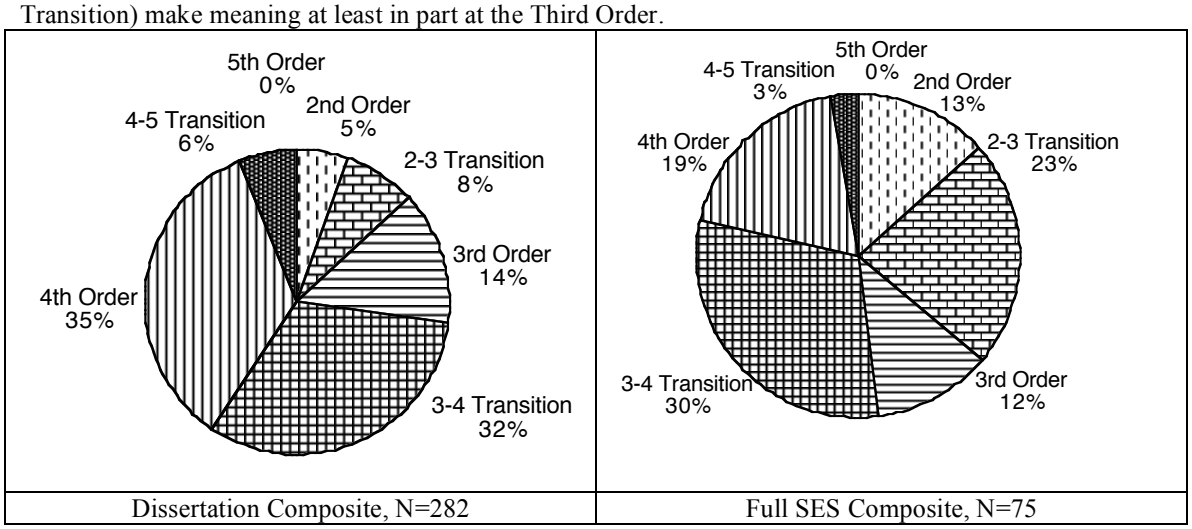

Figure 1: Distribution of Orders of Mind

competing roles or expectations from important external others—unsure how to act, for example, as both a good parent and a good worker if his children and an important project both need time and attention. At the Third Order, people may be "made up by" others' expectations, responding either by cooperating or rebelling, but clearly in reaction to these expectations. Those at the Fourth Order have more options because they have a larger perspective from which to judge, make sense of, and negotiate among expectations. They can identify different internal parts of themselves that may be in conversation with one another, and can take responsibility for their own inner states and emotions—"I feel angry because I *interpret* what you did as a violation of important values of mine, and if I interpreted your actions differently I might feel sad instead." People at the Fourth Order internally mediate among abstractions through abstract systems and ideologies and can have some perspective on culturally or socially mediated definitions of reality. Kegan calls this Order "Self-authoring" in that an individual constructing reality in this way can identify her own role in generating her understanding of the world and is not unduly shaped by the context in which she finds herself. This way of making meaning is also a primary one for most adults (between 3/4 and 1/2, see Figure 1).

Finally, Kegan claims that a small percentage of adults primarily in mid-life or beyond move towards the theoretically posited Fifth Order, where they come to see the Fourth Order's personally created ideologies themselves as constructed objects from a "dialectical" or "self-transformational" perspective. At the Fourth Order, a person can take a perspective on externally imposed values and expectations but cannot see how his own personal system for mediating among these is limited by historical, cultural, psychological, personal, and other forces. Such a perspective on the constructed nature of one's meaning-making system is gained at the Fifth Order. Conceptual frameworks in this view embrace contradiction and paradox. Social relationships are characterized by an integration of self and other—that is, "any aspect of what I used to see as 'my' identity is in part defined by the contrast and relationship with what I used to see as 'yours.'" This perspective is really quite rare, with only a small percentage of people showing any Fifth Order thinking and none being fully Fifth Order in studies thus far (see Figure 1).

In Kegan's framework, development does not occur all of a sudden, though particular incidents can be important catalysts for change. Rather, Kegan argues that people move from fully constructing their understanding in a way that is consistent with a particular Order, towards building a bridge to the next Order by constructing meaning in two ways at the same time, though perhaps preferring one over the other, and eventually towards stepping fully beyond the earlier Order by incorporating it into the larger frame of the later Order. There are four such steps between any two pure Orders—steps which have been instantiated in the subscales of the Subject-Object Interview (Lahey, Souvaine, Kegan, Goodman, & Felix, 1988). Specifically, between any two full Orders—call them X and Y—the four substages are described as X(Y), X/Y, Y/X and Y(X). For example, between 3 and 4 there are 3(4), 3/4, 4/3, and 4(3). In X(Y) there are hints of something beyond X but not yet a fully Y way of making meaning. In X/Y and Y/X both X and Y ways of making meaning are present, with the leading letter, either X or Y, being dominant. In Y(X), there is only a full Y way of making meaning, but the person spends energy defending against slipping back into an X meaning-making structure.

Characteristics of measures

The Subject-Object Interview (SOI) is an unstructured interview which begins by asking the respondent to recall and write personal notes about recent events that capture strong emotions or beliefs—sad, angry, torn, touched, success, strong stand, lost, anxious, changed, or important to me. Respondents choose which situation(s) to describe and the interviewer probes to get at the meaning-making structure behind these beliefs. Interviews are typically one hour long and are audio-taped and transcribed. The transcripts are read carefully and scored to one of the 16 substages between Second and Fifth Order based on the structure of the Order of Mind expressed by the respondent. (A modified interview for children can be used to score the five substages in the First to Second Order transition.) When trained researchers conduct and score interviews, inter-rater reliability to within one substage is greater than 80%.

To make a more efficient measure of constructive-developmental difference we needed to constrain this open-ended format in a variety of ways. First, we decided we wanted some kind of paper-and-pencil measure, rather than an interview. We were encouraged by the work of Baxter-Magolda (1987) who found the written Measure of Epistemological Reflection (MER) was reliable and valid when compared with the open-ended Perry interview. The key to this, she says, is that a "Direct request for justification elicits the essence of the respondent's thinking without restricting the frame of reference" (p. 447).

We also decided to constrain the content realm of the measure, focusing on issues that arise in teaching and professional development. Such a narrowing may raise questions about the domain specificity of our measures and whether they could be used to more generally describe constructive-developmental level. In the SOI, the ability to probe for structure means that no matter what themes people bring up (assuming a reasonably good interviewer), the analyst will get structural content to address. However, this isn't possible in the less interactive format which we are creating. While some theorists propose giving different ratings in different realms (e.g., Fischer & Rose, 2001; Kitchener, 1986), we decided that we're interested in complexity of thinking *within the domain of teaching* and therefore, that there is no problem in only measuring within that domain.

Finally, we decided to limit the precision with which we wanted to measure constructive-developmental differences, focusing only on fully running Orders of Mind and therefore collapsing the 16 substages between Second and Fifth Order into seven—the parentheses substages below and above a full order would just be called that order (so 3(2), 3 and 3(4) would all be called "3"), and we wouldn't distinguish between the two substages in which both orders are running (so 3/4 and 4/3 would both be called "3-4").

Even with these simplifications, creating a more efficient measure of constructive-developmental difference posed several key challenges. Chief among these is that what is being measured is not specific beliefs (the "what" or the "content" of someone's beliefs), but how those beliefs are held (the "structure"). In fact, the same beliefs can be held by people at several different Orders of Mind. For example, teachers at both the Third and

Fourth Orders may hold constructivist views of teaching. However, at the Fourth Order, teachers may have decided on these views because they've experimented and found they like how children respond when given a chance to explore and figure things out together; whereas at the Third Order, they may hold these views because respected colleagues and teachers have vouched for them. Not all beliefs can be held at any Order of Mind—certain kinds of thinking are only possible once someone has reached a particular Order of Mind. For example, one can't really anticipate and consider someone else's thoughts and feelings for their own sake until having reached the Third Order. Still, the *way* beliefs are held can make a difference in how strongly they're held or what it would take to change them. Hammerman (2002) shows how the robustness of beliefs held by those at the Third Order may depend on external social supports for those beliefs, whereas those at the Fourth Order may need to be convinced by arguments to change their minds.

In general, people are much more likely to tell us the content of their beliefs than the underlying way these beliefs are held. When conducting an SOI, learning to probe to elicit constructive-developmental structure is one of the most difficult, but also one of the most important things an interviewer can do. To get at structure rather than content in our written measures, we would have to anticipate likely responses as well as the questions that would probe for structure. Alternatively, we would need to describe or otherwise point towards structural differences and ask people to distinguish among them. In fact, based on these considerations, we've been designing and testing two different measures that we think may serve as alternatives to the SOI. The first is a "Teacher Decision-Making Vignettes" measure and the second, a "Support Beliefs Survey."

The "Teacher Decision-Making Vignettes" measure presents descriptions of four scenarios typical in teaching that pose potential dilemmas for teachers—questions about the value of a teacher-run, collegial inquiry group as follow-up to a professional development workshop; how to deal with a parent angry over too much homework; negotiating among students in disagreement over the "right" answer; and how to respond to a mandated curriculum pacing guide. Teachers are asked to: 1) describe their own responses to questions raised in these scenarios and the reasons for those responses; 2) rate how similar their own way of thinking is with three opinions on the dilemmas posed in the scenarios written from a perspective reflecting 2nd, 3rd, and 4th Orders, and explain their ratings; and 3) pick which of the three perspectives is *most* like their thinking (even if they're all pretty similar or pretty different) and explain why. We hoped that, like Baxter-Magolda's MER, written justifications of thinking would give us structural data. Failing that—or better, in addition to that—we hoped people would be able to recognize and pick reasoning that was structurally more like their own if we presented it to them.

In writing common scenarios to which everyone would respond, rather than beginning, like the SOI, with open-ended generation of actual experiences, we were trying to make it possible to ask generic probing questions that still might be appropriate to the topic at hand. The danger was that these scenarios wouldn't actually feel powerful to respondents; they might not get them close to strongly structurally-linked content. Each scenario was carefully crafted to elicit strong opinions and get at issues that typically differentiate people at different Orders of Mind. For example, we set up a

conflict between two different perspectives and tried to see if teachers could coordinate the two; we set up a tension between internal and external loci of authority to see which was strongest; we tried to assess whether teachers were more concerned about concrete consequences, or role expectations, or self-generated values. Together, we hoped these would give us clues about teachers' Orders of Mind.

In the second measure, the Support Beliefs Survey, we took one step further the strategy of hoping that people would recognize structural differences in thinking. We asked people to rate their agreement with a large number (about 60) items about different kinds of supports for teaching and change in teaching, each written from the perspective of someone at 2nd, 3rd, 4th and to a small extent, 5th Order. For example, an item that was intended to reflect Third Order thinking said, "When I observe my colleagues, I compare my teaching with theirs to see if I am doing my job correctly." A comparable item reflecting Fourth Order thinking said, "When I observe my peers, I clarify my own view of teaching by comparing my view with what I see."

Methods

We've been testing, and continue to test the reliability and validity of these measures in several ways. The Vignettes measure was administered to ten teachers from two Boston area schools who were participating in our NSF-funded project. These teachers, who had signed up to attend a reform-oriented math professional development program during the summer of 2005, had volunteered to participate in our study and were paid a stipend for their participation. In addition to the Vignettes measure, and the initial version of the Support Beliefs Survey, they had been interviewed using the SOI, and had also completed a variety of other measures about such things as their mathematical content knowledge, their beliefs about teaching, and their classroom teaching practice. Analyses of the relationships among these other measures will be reported elsewhere.

One measure of the validity of the Vignette measure was teachers' reactions to the scenarios. Some teachers found the Vignette scenarios very realistic and thought provoking in their own right. One said he thought these should be used in teacher education courses as dilemmas for novice teachers to ponder. Thus, in a general way, the content and issues we included in these dilemmas was relevant to teachers' practice. As another validity test for the Vignettes measure, we conducted cognitive process interviews with teachers, asking them to walk us through their thinking about various Vignette items so that we could try to understand whether their thinking about the Vignettes measure was responding to what we intended. We coded responses to the vignettes themselves based on Kegan's theory as if they were an SOI transcript, but also using grounded theory methods (Glaser & Strauss, 1967) to look for other themes or issues arising in the responses. These included knowledge and beliefs about mathematics, beliefs about math teaching and learning, Myers-Briggs personality characteristics such as sensing-intuiting, or relational (feeling)-analytic (thinking), as well as question quality categories such as content distractors, language confusion, and the like. We also looked for correlations between people's actual SOI scores and their ratings of agreement with the 2nd, 3rd, and 4th Order perspectives. In the results section, we'll report on some of

what teachers wrote and said in our initial draft of the measure, and how we've used this to re-shape and revise the measure.

The Support Beliefs Survey items were also originally tested with the 10 teachers described above. We wanted to be sure that the statements we had generated validly reflected different Orders of Mind, so we asked a few reliable SOI scorers to say how they would most likely rate these statements if they were said by someone in the context of an SOI; to describe a range of plausible levels that they might score the statement in an SOI; and to describe possible confounding variables. We used this feedback to rewrite the items, modifying them to limit confounding variables, focusing items that seemed to be about several things, and deleting items that seemed too difficult to fix. We then sent the new set of 59 items out to a larger pool of experts for confirmation of validity again asking about most likely SOI score, plausible range of scores, and possible confounding issues. At the same time, we distributed the survey itself quite widely to teachers and prospective teachers using an online survey administration tool, Survey Monkey (http://www.surveymonkey.com), and have thus far received over 250 complete sets of responses. We plan to analyze the pattern of these responses using confirmatory factor analysis (CFA), with the input from our experts serving as the basis for the hypothesized structures we will test (Kline, 2005). In the results section, below, we will describe some of our methods for refining the survey, although the confirmatory factor analysis is not yet complete.

Results

The Vignettes measure, though not an interview, is still an intensive measure requiring a substantial amount of time to complete and to score. The Support Beliefs Survey takes just 15 minutes to complete and will be scored by applying factor loadings to generate a maximum likelihood estimate of SOI level. If the Support Beliefs Survey can be shown to be a reliable and valid measure of constructive-developmental differences, it would certainly serve as an easy-to-use tool for large-scale research. However, on its face there is reason to doubt that merely recognizing and agreeing with a set of statements accurately reflects constructive-developmental level. Still, if our confirmatory factor analysis shows that underlying factors representing constructive-developmental categories (as judged by expert SOI scorers) accurately predict these responses, then it may serve as a rough proxy for the SOI.

Responses from the Vignette measure, on the other hand, require actually generating responses which are more likely to reflect structure. At this point, however, we only have responses to the Vignette measure from the ten intensively-studied teachers. Unfortunately, their range of SOI levels is somewhat limited, with all teachers but one scoring in the range 3/4 to 4—which is to say, with a 4th order structure fully running and, for the most part, dominant. This means that we don't have enough data from people at varied developmental levels to tease out how different people respond to our 2nd, 3rd, and 4th Order vignettes. Nonetheless, we can learn important things by looking at our process of analyzing and refining these measures.

In the sections that follow, we give examples of some of the issues we were facing in developing these measures. These include some successes but, for the most part, these issues raise questions about whether the measures can become robust enough to use on a larger scale. We begin by addressing issues associated with the Vignettes measure, and then turn to the Beliefs Survey.

Vignettes measure

Finding Structure in Vignette Responses. In some responses to the vignettes, we were able to find evidence of constructive-developmental structure. In conducting an SOI, interviewers are typically looking for both "floors"—statements which could only be made if an interviewee has *at least* certain constructive-developmental capacities—and "ceilings"—repeated chances for interviewees to demonstrate a higher Order of Mind which they seem unable to take. Scoring an SOI involves narrowing the range between floors and ceilings until ideally a single substage is identified. Our experience with SOIs suggests that floors are much easier to find than ceilings. This was also true, even exaggerated a bit because we couldn't offer opportunities for testing upper limits, in looking for evidence of Order of Mind in the vignette responses.

Thus, there are many places in the vignettes where respondents exhibited at least the capacities of a Third Order of Mind. Participants were often clearly able to take the perspective of a colleague, parent, or student and look at issues abstractly—both Third Order capabilities. For example, one teacher wrote, *"Fairness is important to children—involving them in the discussion so that they see and understand different thinking would be important."* Her concerns are about abstract values such as "fairness" and ways to help students see and understand different perspectives, as opposed to getting right answers. In the same vignette about a student disagreement over the "right" answer, another teacher wrote in reaction to a Second Order response, "*It's not a matter of what's right or wrong but rather a matter of interpretation. Some students are ready to see different points of view, others are not.*" This teacher is beyond a dualistic view of knowledge—a characteristic of Second Order thinking. She has sufficient complexity to see and reflect upon others' points of view—a characteristic of the Third Order. Yet another teacher wrote about the same vignette, "*Life has as much 'gray' as it does 'black and white' answers. Our role is not to shape students into little robots but to help them expand their thinking and be open to new ideas*." This teacher was able to consider different roles she might play in the growth and development of students, demonstrating at least a Third Order view and even some consideration and choice among roles, suggesting some Fourth Order perspective as well. In the professional development vignette, the ability to take perspectives that differentiates a Third from a Second Order person was evident in this participant's praise of the described workshop. "*We could see what students experience mathematically before and after our grade level. I think it's good to walk through the math as a learner first...It's important to know how the kids feel as they learn*." In fact, coordinating the multiple perspectives which constitute a learning trajectory may indicate a bit of Fourth Order thinking.

There were other instances where teachers gave evidence of thinking that seemed to reflect an Order of Mind similar to what their SOI revealed. When asked in the

professional development vignette how a colleague's thinking about a new curriculum might affect her own, a teacher scoring at the Fourth Order wrote, "*Her thinking would not affect how I was viewing the curriculum. I'd need to try out things for myself*." This teacher is not embedded in the external authority of her colleague. In fact, she clearly wants to make up her own mind about how the curriculum works for her—a characteristic of a Fourth Order, self-authoring perspective. Her response to a colleague who complains in the pacing guide vignette shows that she can coordinate her perspective of the guide with her perspective of a teacher's professional judgment. "*Her years of experience are valuable. I'd ask her if she couldn't incorporate the guide to some extent without compromising the sound math education of the kids she feels she's capable of providing*." About the same vignette, a teacher whose SOI score was 3/4 (Third Order dominant but with a Fourth Order structure fully running) says:

> *This policy would focus my content, make me plan more carefully. I would stay on track…This reminds me of* [how] *successful my math program was these past two years. I teamed with the other multiage teacher. She took my first graders, I took her second graders, and we only did one grade level each. This structure caused us to commit to an hour a day of math and we were faithful to this hour, no matter what, because it would mess up the other class if we cancelled. Thus, math time was sacred, not to be interrupted.*

This teacher seems to need someone else, or a role expectation, acting as an external authority to help her keep her math time "sacred." At the same time, she can be part of creating this structure collaboratively with her colleague, which then serves this higher purpose. These two perspectives together seem to point towards both Third and Fourth Order views, confirming what we know about her from the SOI.

Sometimes there was implicit structural content in teachers' reactions to the Second, Third and Fourth Order responses in the vignettes. This is because the logic of earlier Orders sometimes seems immature to those at more sophisticated Orders. We wanted the particular views expressed in the responses to be the same, so that only the reasoning behind what was going on differed. For example, in the pacing guide vignette we decided to make each of the three responses in favor of the guide and supportive of the coach who is charged with being sure teachers follow the guide exactly, but for different reasons. The Second Order response was a fan of the guide because it was easy and people who don't do what they are told to should get into trouble. This teacher didn't mind the coach coming in because once s/he started following the guide exactly, there were no more complaints. The Third Order response was a fan of the guide because it built on National Standards, because experienced respected colleagues, including the math coach, supported it, and because having a guide made role expectations clear. The Fourth Order response was a fan of the guide because, on reflection, it just codified her/his own existing beliefs about good teaching, because it served as a resource, and because talks with the math coach helped clarify her/ his thinking and generate new ideas.

Although all these views of the guide and the coach were positive from within their own perspective, people perceived the Second Order view of the coach as authoritarian because it didn't question the enforcement of the rules. For example, one teacher said,

"*This coach doesn't appear to be a coach at all, just a dictator*." Another teacher makes clear that she would change her mind and actions based on being convinced, rather than responding to concrete consequences. "*I certainly disagree with the use of disciplinary action. A far more effective approach would be to demonstrate the advantages and benefits of trying the pacing guide.*" The participants also seemed to take issue with the Second Order responder because she did not consider anyone's needs but her own. As one teacher wrote, "*Job may be easier, coach may be happy, but are the children learning and happy with their learning?"* We feel these participants maybe be reacting against the concreteness and lack of perspective-taking a Second Order adult would exhibit. This would indicate that our participants are beyond Second Order themselves, something confirmed by our SOI analyses. These teachers also sometimes seemed to recognize how others might see these situations even if it was different from their own, confirming that they have at least the Third Order capacity to do so. For example, a teacher who really wanted to be responsive to the varied individual needs of students also knew that "*Some teachers may like it because it makes planning easier.*"

More participants agreed with at least parts of the Third Order response of the same vignette. Many of these agreed for reasons that seemed more about tone or content, "*This teacher seems more relaxed,*" for example. However, a teacher in the 3-4 transition wrote, "*I would also feel conflicted if someone (colleague) I respected disagrees (re: guide) with someone else I also recognize as knowledgeable on same subject (coach)."* Another teacher whose SOI score showed her to be dominant Third Order wrote, *"I think it's insubordinate to oppose the math coach's direction and I don't support the experienced teacher's complaints*." These two comments suggest Third Order thinking, though it is difficult to tell without being able to probe, as we will describe in the next section. The ones that did not agree seemed to react against the Third Order views of the response. One teacher wrote "*I would not let my colleague affect me at all*" which may reflect a lack of embeddedness in the colleague's authority. Three other participants seemed to notice that this responder was embedded in the authority of the coach. One such response was "*this teacher feels like she is accountable to the math coach to follow the guide rather than her students to teach them the math they need."*

Reactions to the Fourth Order response tended to be positive across the board. This may have been due, in part, to issues of content inadvertently written into the vignette—this teacher seemed more at ease because she was not having to change her practice to meet the mandate of the guide as the guide already aligned with her teaching. Therefore, it's difficult to tell if the participants were picking up on structural differences, or what seemed to be the ease and confidence of this teacher. Many respondents wrote about ways of thinking that seemed to have a Fourth Order, "self-authoring" flavor. "*The key line for me is: '…since it aligns with my own standards for good teaching.' It would feel very different if I was asked to implement a program I strongly disagreed with.*" But not all respondents wrote from this perspective. One of the two participants whose SOI was scored with Third Order dominant sounded like she still saw the guide as a source of authority. She wrote, "*After reading over the guide, it confirms that what I'm doing is aligned with the curriculum guide, after all…This person has a positive and a productive attitude about the directive from the district."* Still, we are unsure whether participants preferred this response in greater numbers because they were mostly Fourth Order

themselves. Teachers at the Third Order might be attracted by the open-mindedness and more student-centered view as their schools are promoting differentiated instruction; but their explanations might still reflect a Third Order focus on roles and an external authority, in contrast to those at Fourth Order.

Next, we will describe some of the difficulties we faced in designing a vignette measure as an alternative to the SOI.

Reduced Confidence Due to Inability To Probe. As noted above, we were struggling with how to probe for constructive-developmental structure given the written response format of this measure. In the Vignette measure, we asked what participants thought about certain scenarios and why. Although we were hoping the "why" might give us information about participants' meaning-making, many of the answers we received were based largely on content. For example, for the vignette about a pacing guide, one participant stated, "*A planning guide that dictates which page to assign daily does not appeal to me - it does not take into consideration children's individual differences.*" This statement doesn't help us differentiate Order of Mind—a teacher at 4th Order might say it based on a self-generated value about the individual learning needs of her students. However, a teacher at the 3rd Order might say the same thing, but basing the statement on fitting in with the culture and beliefs of her school or colleagues who support differentiated instruction. During an SOI, we can ask follow-up questions such as, "How do you know that taking account of children's individual differences is the right thing to do?" to see if the source of this belief is internal or external. On a questionnaire, however, we cannot tailor questions to an individual's response, which limits what we can know with confidence from these responses.

Even when we found an answer that sounded structural, our inability to probe meant that we were making assumptions about the reasons behind participants' comments and couldn't get at added complexity. One teacher, whose SOI score was 3/4 (meaning both a Third and Fourth Order structure, with Third Order dominant), supported the pacing guide for reasons that seemed to reflect a 3rd Order need for people to fulfill externally defined roles. "*I think it's insubordinate to oppose the math coach's direction and I don't support the experienced teacher's complaints. It sends a confusing message to peers."* During our follow-up interview, we learned that she had been a math lead teacher when her school adopted the *Investigations* elementary math curriculum, and had been frustrated when administrators were lax in enforcing new curriculum guidelines with her teacher-peers. *"There's always manipulation in cozying up, and sometimes administrators look the other way. There often are different rules for different teachers. This then undermines the goal of the pacing guide when teachers view inconsistently the rules and people are lax themselves in following them."* She complained about teachers not fully engaging in the curriculum adoption and implementation process. "*One person showed up at 10, and she didn't come all the time, and then when we had our meeting about deciding what we were going to do for the next year, she was complaining about the program. And I wanted to say, 'How come you get to complain when you didn't participate?'*" This seemed especially important in this vignette because, as she said, "*I think I was interpreting that this was a program I believed in…If the coach said you have to use the basal reader, then I might be insubordinate*." We wondered, were these

interview comments confirmation of her holding a Third Order view that sees the world in terms of prescribed roles and is very upset when they're violated? or did the fact that her complaints depended on her opinion of what was being implemented show some perspective on how those roles and the rules were intended to serve larger purposes? In fact, her SOI score of 3/4 would suggest that she makes meaning in both of these ways. However, her response on the vignette was not as nuanced. Where hypotheses about individuals could be confirmed or disproven during a well-conducted interview, we had to remain much less confident with our written measure.

Issues with Pre-Set Scenarios. Although we hoped that providing a scenario would create a common starting point to which all respondents could relate, we faced several issues that make the future use of such scenarios uncertain. First, we could not possibly provide enough content information so as to completely describe the scene, so participants were sometimes left feeling uncertain about what was really going on. One teacher said in the follow-up interview,

> *Some of it was difficult because I wished I could have asked questions or seen what was going on, and it somewhat seemed hard to interpret because you weren't actually there, or you didn't hear the feedback or the discussion between the people. You just kind of had to assume certain things.*

Other times the lack of detail led participants to fill in the blanks with information from their own experiences. For example, in one participant's mind the professional development workshop described in the vignette was similar enough to the professional development workshop they were participating in, that she responded to the vignette as if it were talking about the same workshop. Another participant imagined that the pacing guide was about a curriculum she liked, since the curriculum wasn't specified.

When the content we provided did not match the experience or beliefs of participants, they also tended to have difficulty. One teacher was very clear on how difficult it was to put herself into the place of someone with very different beliefs. "*It just didn't seem like there was room—I guess what I didn't like was the fact that you have to take this point of view, because that would not have been my point of view…I mean, the decision has been made that…they're going to do the homework and that's pretty much the bottom line*." Since she would have made a different decision about *what* to do, it was difficult to think about the different structural *ways* of thinking about this issue.

Another participant told us, "*I think you just didn't know how to approach it…some of the attitudes were so foreign to me. I had a lot of trouble saying—well, I wouldn't do anything close to what they're doing, and so I think they* [the vignettes] *were really hard to do*." A third participant claimed that the situation described in the vignettes was unrealistic to her school situation, "*The other thing I found difficult was that there's parts of them that are realistic in terms of our school district and our thinking, and there's parts of them that are not.*" For example, the homework vignette seemed unrealistic to her, since her schools' homework policy was clearly communicated and, "*So to have somebody to the point of needing to go to the principal to complain about what you're doing or what you're not doing has escalated to a point that it should not have.*" We are

uncertain how to deal with these difficulties caused by people holding different beliefs from those described in the scenarios.

Content as a Confounder. Although our intention was to focus attention on structure rather than content, as we begin to see above, it was impossible to avoid specific content in crafting the scenarios, and this often confounded teachers' responses. Teachers tended to disagree with the idea of a pacing guide because they thought they should have the authority to modify curricula to meet the needs of individual students. "*I currently adjust day-to-day curriculum based on individual student needs. I am free to spend as much time as needed for students to demonstrate concept mastery. A pacing guide would override my professional judgment about the curriculum priorities and needs*." Other teachers objected to the goal of the guide and the coach's role, which seemed to be about making things easier for teachers, not focusing on what students need. "*There was no mention of the students in this 'train of thought.' It sounds like this teacher is only thinking about how the change will impact him/her*." Thus, the way we presented the pacing guide scenario seemed to challenge teachers' views about their professional authority, and their values about focusing on students' needs rather than the teacher's. They responded from this perspective, rather than based on differences in the structure of the responses.

Thus, we've had both some success and some difficulties in using the Teacher Decision-Making Vignettes measure as an alternative to the Subject-Object Interview. If we had been able to test the measure on teachers with a wider range of constructive-developmental levels, we might have seen more obvious differences in responses to the several measures. As it was, we saw only minimal differences that could be attributed to SOI level—instead, the content of the beliefs often seemed more salient. However, because the measure is so time-consuming to administer, it wasn't practical to test it with a larger group. The Support Beliefs Survey *could* easily be tested with a large group, though, although it had some different difficulties. We turn to those advantages and difficulties next.

Support Beliefs Survey

We began designing the Support Beliefs Survey to understand directly teachers' views about what they would consider supportive in making reform-oriented changes in practice. Because we brought a constructive-developmental theoretical perspective to the question, we designed items that would reflect the concerns and orientations of people at different Orders of Mind. Soon, we began to think that people's affinity for items reflecting these different perspectives might serve as a proxy for Order of Mind, and began designing the measure with this goal.

It is difficult to capture important aspects of Order of Mind in brief statements about supports. Although the statements we generated are based on our experiences reading and scoring a variety of Subject-Object Interviews, and we tried to represent some of the diverse ways that people can hold their beliefs at each Order of Mind still, to some extent, these statements were caricatures of important features of the different developmental levels. So, for example, the items that we intended to reflect Second Order thinking

depicted people as focusing on ease of use of an innovation, not getting into trouble, doing things because of concrete rewards, and having difficulty really considering others' perspectives. It's not that these characteristics don't accurately reflect this theoretical position, it's just that actual people are more complex, for example, looking for one or another of these in different circumstances.

Statements intended to reflect the Third Order depicted people looking to outside experts, more experienced colleagues, or people in positions of authority to describe best practice or the right way to do things; a concern about fulfilling externally defined roles and being seen to be doing well; and difficulty coping with conflicts among these views. However, we also had one or two examples of a reactive Third Order orientation—one which is defined by external expectations but chooses to do exactly the opposite, like a rebellious teenager always doing just as his parents tell him not to.

Our depiction of Fourth Order was of someone with an internal sense of authority, coordinating multiple perspectives, and maybe being just a little too attached to their own way of deciding. But still, in total, the caricature of Fourth Order may have been a little over-confident and too sure of themselves. Items that contained a bit of a Fifth Order perspective suggested questioning these self-created systems and embracing an integration of multiple perspectives. We tried to tap not only the essential orientation of each Order, but also some of the essential limitations—that is, the lack of capacity to take or coordinate perspectives or values characteristic of each Order.

We wanted to write items that might tap different ways of experiencing the different Orders of Mind. For example, as noted above, we wrote items representing both conforming and reactive versions of a Third Order view. However, having done so, analyzing these may pose some difficulties, as people at the same Order of Mind would respond very differently to items that are supposed to tap that Order.

More subtly but along a similar vein, it's become clear from some feedback offered by a few participants in the larger survey that whether one accepts the advice of an expert or principal or more experienced colleague may depend on whether one respects their expertise in the particular area of concern. For people at the Fourth Order, this question of respect is a matter of weighing what they know and believe and how that fits with the values and goals of the person him/herself. For those at the Third Order, it's a question of who constitutes the "culture of embeddedness" in which the person lives—that is, who represents "authority" to them. Respondents might answer questions differently depending on whether their experiences would support giving respect to the experts named in the items—a difference which may complicate our analyses.

Similar issues arose concerning other variations we tried to incorporate into the items—a relational vs. independent orientation, varying degrees of experience, MBTI feeling vs. thinking and sensing vs. intuiting styles, and so forth. Even after our revisions, we may have left in a confounding of conflict avoidance with lower Orders of Mind. We will try to test this in our confirmatory factor analysis.

Finally, our revisions tried to clarify and focus the questions from our first draft of the survey. When we looked carefully at the questions, some seemed to be making two statements and we worried that people might respond in two different ways at once and therefore we wouldn't know how to interpret their rating. We focused others by bringing together different goals into a single feeling of tension or conflict; or by clarifying the relationship of respect to people in the statement. We also made some statements less absolute so they'd appeal to a wider range of respondents. Some examples of these changes are listed in the table, below.

First draft	Revised version	Changes
I believe there's just one best way to teach and that experienced teachers and experts can tell us what that is.	I believe experienced teachers and experts can tell us what's the best way to teach.	Was two statements, now one
I prefer to set my own learning goals and then seek out professional development activities to help me meet them.	I prefer to set my own learning goals for professional development.	Was two statements, now one
I have a sense of my own values about teaching and learning, but I also want to do what others expect me to, even if their expectations differ from my own.	When my own values about teaching and learning differ from what others expect of me, I feel torn or confused.	Was two statements now one, Focus on unitary feeling rather than conflicting goals
I would feel excited participating in a group where teachers explore plausible explanations and arguments and respectfully critique each other's ideas.	Participating in a group where teachers explore plausible explanations and arguments and respectfully critique each other's ideas would help me teach better.	Shift focus from feelings to improved teaching
I only really participate in professional development workshops because I get paid to do so.	Getting paid or receiving classroom materials is the main reason why I participate in professional development.	Less absolute
I get uncomfortable if people disagree when they talk about best practices.	I get anxious or confused when people I trust disagree about best practices.	Clarify relationship

Thus, many of the issues that arise in the context of the Vignettes measure also come up with the Support Beliefs Survey. Although the measure may be far from perfect, we believe that it's getting clearer and better and is ready to test on a large scale.

Discussion

Our experiences creating alternative measures of constructive-developmental difference/ Order of Mind point to some of the difficulties of trying to capture the structure of people's meaning making through an easily administered instrument. Thus far, we have had mixed success in generating responses that indicate structural differences. With the vignettes measure, this was due in large part to the limited constructive-developmental variability and small size of our pilot sample. We are hoping that the large-scale sampling we've done for the Support Beliefs Survey will yield a more variable group.

While we find some evidence of constructive-developmental Order of Mind in statements made on both instruments, this is often mixed with responses reflecting specific content beliefs. This is complicated further by respondents' tendency to either read their own circumstances into the scenarios, or to resist doing so and therefore dismiss the scenarios as unrealistic. That is, some of the assumptions that we had to make

in writing the items didn't always fit people's own assumptions and circumstances, and this muddied their responses. In addition, as we tried to depict the several different ways that someone at each order would think, we may have created multiple response sets that all point towards the same Order. Further analyses will be needed to determine whether this helped us uncover structural differences, or confused the situation.

We also described some methodological issues that arise in trying to create such instruments. Chief among these is the inability to responsively probe for underlying structural differences. Our guess is that this will mean we are more likely to observe "floors" than "ceilings" in the Vignettes measure. We are less certain what the implications are in the Support Beliefs measure, but it will certainly reduce precision of the instrument. The further analyses we plan to conduct in the coming months will help determine whether these problems and issues can be overcome.

Finally, there are a number of issues that must be addressed even if we are successful in creating these alternative measures. Chief among these are ethical questions about how such instruments would be used. Measurement tools are often seen as precise and certain and, therefore, the information they convey is often used to make fine grained distinctions between people, often inappropriately. The measures we're creating, like many others are unlikely to be very precise, and we don't want them mistakenly used as if they were. For example, although we think attention to constructive-developmental differences can have an influence on the design and implementation of teacher professional development programs, we don't mean these measures to be used to select teachers for one type of program or another catering to those at one Order of Mind or another. Instead, we hope that these tools will be used for research to help providers of professional development understand differences in how teachers experience the same program, and to improve the design of programs so that they will support the varied learning approaches and needs of a wide range of teachers. Our image is that understanding differences in what learners—whether teachers or students in classrooms—bring to a learning opportunity will enhance and improve the effectiveness of what educators—whether teacher educators or teachers—do. This research is intended to support understanding of one kind of such differences.

References

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: A handbook of policy and practice* (pp. 3-32). San Francisco: Jossey-Bass.

Ball, D. L., & Rundquist, S. S. (1993). Collaboration as a context for joining teacher learning with learning about teaching. In D. K. Cohen, M. W. McLaughlin & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Baxter Magolda, M. B. (1987). Comparing open-ended interviews and standardized measures of intellectual development. *Journal of College Student Personnel, 25*(5), 443-448.

Borko, H., & Putnam, R. T. (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. R. Guskey & M. Huberman (Eds.), Professional development in education: New paradigms & practices (pp. 35-65). New York: Teachers College Press.

Brahier, D., & Schaffner, M. (2004). The effects of a study-group process on the implementation of reform in mathematics education. *School Science and Mathematics, 104*(4).

Fischer, K. W., & Rose, L. T. (2001). Webs of skill: How students learn. *Educational Leadership*, 6-12.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine Publishing.

Hammerman, J. K. (2002). *Experiencing professional development: A constructive-developmental exploration of teachers' experiences in a mathematics teacher professional development program.* Unpublished Doctoral Dissertation, Harvard Graduate School of Education, Cambridge, MA.

Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127-150). San Francisco: Jossey-Bass.

Kegan, R. (1982). *The evolving self: Problem and process in human development*. Cambridge, MA: Harvard University Press.

Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.

Kent, A. M. (2004). Improving teacher quality through professional development. *Educational Leadership, 124*(3), 427-434.

Kitchener, K. S. (1986). The Reflective Judgment Model: Characteristics, evidence, and measurement. In R. A. Mines & K. S. Kitchener (Eds.), *Adult cognitive development: Methods and models* (pp. 76-91). New York: Praeger.

Kline, R. B. (2005). *Principles and practice of Structural Equation Modeling* (2nd ed.). New York: The Guilford Press.

Lahey, L., Souvaine, E., Kegan, R., Goodman, R., & Felix, S. (1988). *A Guide to the subject-object interview: Its administration and interpretation*. Cambridge, MA: Harvard University, Graduate School of Education, Laboratory of Human Development.

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. *Phi Delta Kappan, 76*(8), 591-596.

Mundry, S., & Dunne, K. (2003). *Teachers as learners: Professional development in science and mathematics—Facilitator's guide*. Thousand Oaks, CA: Sage.