Understanding What Children Know About History:

Exploring the Representation and Testing Dilemmas

S. G. Grant
University at Buffalo

Abstract

Understanding what children know about history or social studies has proven illusive. In this think piece, I explore two dilemmas—the representation dilemma and the testing dilemma—that surround the question, "How do we know what children know?" I conclude that teachers, researchers, and policymakers must engage in conversations that put students' representations of their historical knowledge and understanding at the forefront.

Introduction

Understanding what children know about history or social studies[i] has proven illusive. Over time, educators and psychometricians have constructed many ways of assessing what children know under the assumption that there is no *one* best assessment. Offering multiple assessments, however, fails to solve the problem of deciding what children's representations on these assessments mean.

One problem is that if children are asked different kinds of questions about the same topic, their responses suggest that they know different things (Dickinson & Lee, 1984; Levstik & Barton, 1997; Nuthall & Alton-Lee, 1995; Rogers, & Stevenson, 1988; VanSledright, Kelly, & Meuwissen, 2006). A second problem is that children may appear to know different things over a relatively short period of time (Rogers & Stevenson, 1988; VanSledright, Kelly, & Meuwissen, 2006). A third problem is the challenge of gaining consensus across evaluators; different scorers may evaluate the same student's response very differently (Baker, 1994). And finally, students' scores on tests, whether classroom or standardized, may be variously interpreted (Horn, 2003, 2006). Taken together, the research literature suggests that both teachers and researchers struggle to answer the question, "How do we know what kids know?"

Such a question raises issues of epistemology, semiotics, constructivism, and the like. These perspectives can be useful, but I come at the question on a more practical level with three reference points in mind: teachers, researchers, and policymakers. These actors, and others, form loosely-knit *communities of practice* [ii] (Lave, & Wenger, 1991; Wenger, 1999), groups that

presumably hold similar values about the kinds of knowledge students ought to hold, how that knowledge ought to be assessed, and how the representations that students offer ought to be judged. These different communities *may* share some common values and assumptions around these issues, but, at this point, those commonalities and any emergent differences have not been identified and examined. Consequently, teacher, researcher, and policymaker communities may be asking the same questions, but their interpretations of the answers may be profoundly different.[iii]

The question of how do *we,* as adults, know what children know has been vexing me for some time. In this think piece, I use interview data with a third grader to explore two dilemmas—the *representation dilemma* and the *testing dilemma*—that surround the question of how do we know what children know.[iv] I conclude that teachers, researchers, and policymakers must engage in conversations that put students' representations of their historical knowledge and understanding at the forefront.[v]

Dilemmas and Conversations: How Do We Know What Children Know?

In the sections that follow, I first describe the nature of what I term the *representation dilemma,* drawing from the data collected on children's ideas about Columbus's interactions with Native peoples. I then turn to a discussion of the *testing dilemma*, wherein I detail five problems inherent in assessing children's historical knowledge and understanding. Finally, I offer a challenge to teacher, researcher, and policymaker communities to engage in a conversation around these dilemmas and the potential for a set of agreed-upon assessment markers.

The Representation Dilemma

Researchers continue to document the challenges to understanding what children know and understand about history (Barton & Levstik, 1998; Lee & Shemilt, 2004; Nuthall & Alton-Lee, 1995; Rogers & Stevenson, 1988; VanSledright, 2002; VanSledright, Kelly, & Meuwissen, 2006). Less clear is to what extent teachers and policymakers understand the instability of students' representations and, more importantly, the consequences of this finding for assessment.

The fluidity of students' ideas is one part of the *representation dilemma*. Fluidity, in some observers' eyes, means that students' representations of their knowledge and understanding are, at best, inconsistent and, at worst, incoherent. Neither of these interpretations is helpful, but they do point to a second element of the dilemma, namely, that what students know and understand is generally represented through their words, whether oral or written. With no means by which to unlock students' *true* historical grasp, we are left with only their oral or written traces, and language is a slippery medium (see, for example, Austin, 1975; Fish, 1980; Searle, 1989).

The third element of the representation dilemma is the context in which students are asked to represent their historical knowledge and understanding. Researchers have a luxury that teachers and policymakers do not. In the context of their work, researchers can puzzle over how to make sense of students' ideas. The consequence of their efforts—insights into what, how, and why students engage (or not) with history—enrich the possibilities for understandings of teaching and learning, in general, and teaching and learning history, in particular. Teachers and policymakers can benefit from such insights, but they bear a responsibility beyond that their researcher colleagues: They must evaluate and assign judgments to students' representations.

Teachers and policymakers may be interested in understanding students' ideas from a researcher's perspective, but they have purposes—evaluation and judgment—that researchers can avoid. Knowing that students' knowledge and understanding is unstable and can only be represented by the medium of language is problematic in its own right and does not obviate the expectation that teachers and policymakers can and will assign a value to the historical representations students offer.

An Illustration of the Representation Dilemma

To illustrate the nature of the representation dilemma, I offer an example of one child's representations of the interactions between Columbus and Native Americans. I interviewed Bill, an urban, African American third grader as part of a larger study that explored (a) questions about the students' knowledge of the relationships between Columbus and Native groups and (b) questions about how students read and interpret two short trade book passages covering Columbus's initial return trip to Spain (Adler, 1991; Krensky, 1991). The first selection says, "In March, 1493, Christopher Columbus sailed back to Spain with gold trinkets, parrots, and a few Indians" (Adler, 1991, p. 22). The second text says, "Soon the *Nina* and the *Pinta* are ready to sail back to Spain. The ships are already loaded with many kinds of food….Columbus has also forced six Indians to come with him" (Krensky, 1991, p. 41). Throughout the interview, I probed for the influences on Bill's knowledge of Columbus and Native populations, for his understanding of the claim that Columbus "forced" some Natives to go with him (Krensky, 1991), and his views about Columbus Day celebrations.[vi]

Bill demonstrates some early factual confusion, but then he rallies to offer a nuanced interpretation that includes the perspective that he and Columbus might not be all that different. Bill's first ideas about Columbus are a jumble of facts, conflations, and wild suppositions. The following are my questions and comments (in italics) which elicited Bill's responses:

> [*Have you ever heard of Columbus?*] We studied him because when he was in his land, he thought, everybody else thought it was, a flat, um, land, so he, he was brave, so he went in his boat and went and bumped into North America and then he thought he, he went to um, India, India, so he called the Iroquoians the Indians.
> [*Were the Iroquoians in India?*] They were in North America.
> [*Oh?*] That was their land and then here comes, um…Europe and they try to go to there and buy the place, they like be trading with [unclear]. They didn't know if, if they were letting them borrow their stuff, but, and they thought, now the land is ours, so they started a war.
> [*Who started the war?*] Um…the, the French, um…the, the, Europe, the Europeans.
> [*Did Columbus start the war too?*] Uh uh, he didn't, he didn't start anything. He was friendly.

Here, Bill represents his ideas in a loose narrative that moves into and out of the factual realm. With this crazy quilt of ideas, it is difficult to know with any certainty what Bill understands about this period or these actors. After listening to the Adler and Krensky texts, however, Bill offers a tentative interpretation:

> [*Some people say that Christopher Columbus is a hero, what do you think about that?*]
> He is a hero, because he found new land and ways they could get food.

My question prompts an interpretive response; instead of soliciting factual information, I am asking Bill to tell me how he understands Columbus in a larger context. He asserts that Columbus is a hero is based on a general sense of the Columbus story ("he found new land") and a literal interpretation of the Krensky text ("they could get food").

As the interview continues, however, Bill's responses suggest that he has an appreciation for the differing perspectives on Columbus *and* an empathetic view that he and Columbus may share some similar traits:

> [*I was thinking about the piece I read that said that Columbus forced six Indians to come with him. What do you think about that?*] That he was real mean to them because they might not have wanted to go with him so he had to force out…anger.
> [*And what does that make you think?*] That he was a little mean.
> [*A little mean? What does that sound like to you?*] Like he's, um…half good, half bad.
> [*Have you ever heard of anybody like that?*] Me.
> [(Interviewer laughs) *You!*] (nodding) My brother and sister, everybody I know.
> [*How is that so?*] Because sometimes they get mean, and sometimes they be good.
> [*Can you be a good person and still do some mean things?*] I do some good things, but mostly all bad. (Interviewer and student laugh).

On one reading, Bill's interpretation of the interaction between Columbus and Native groups is thinly represented. He seizes on the word "forced" from Krensky's text, though only after being prompted, and only then does he attribute harshness to Columbus's actions.[vii] Some of his comments suggest that he sees competing perspectives on this event—i.e., those of the Natives, of Columbus, and earlier, of Europeans in general. In no case, however, does Bill explicitly draw on factual evidence to support and enhance his claims.

What stood out to me, however, was Bill's ability to see Columbus as a fully-figured human being. As a historical actor, Columbus is neither unfailing hero nor devil incarnate. Bill recognizes that Columbus added to *and* subtracted from the human condition. And in doing so, Bill holds history up as a mirror: Columbus is "half good, half bad" just as he, his siblings, and "everybody I know," are. In these seemingly simple phrases, Bill connects past and present on a human level. He is neither judging Columbus by some naive measure of political correctness, nor is he misapplying the standards of the present to the past.[viii] Instead, Bill judges some of Columbus's actions as "a little mean" because he knows some of his behavior can be judged similarly.

Interpreting the Representation Dilemma

A teacher reading Bill's words might be dismayed at his weak factual knowledge but take heart from his ability to rally to a plausible interpretation of and personal connection to Columbus. A researcher, however, might focus on the interaction among Bill's prior content knowledge, his lived experience, the texts read, and my interview questions, and conclude that Bill demonstrates a weak ability to reason from evidence. And a policymaker might find the 25-

minute interview interesting, but conclude that, if tested on an objective-style assessment, Bill is likely to perform poorly.

As I thought about my conversation with Bill and the other students in the data set, I identified three components of what appeared to be a *representation dilemma*: (a) the fluidity of students' historical representations; (b) the slippery nature of language, and (c) the need to evaluate students' representations.

The fluidity of students' historical representations. As an example of this problem, consider three points within Bill's response—before the Adler and Krensky readings, immediately after the readings, and after my question about Columbus forcing Natives to return with him to Spain.

In response to my initial questions, Bill demonstrates a weak factual base: Columbus encounters "Iroquoians" in "North America," while proving that the earth is not "flat" and creating "friendly" relations with the Natives. His narrative offers some accurate information (e.g., Columbus believing that he had landed in the Indies), but overall, his account resembles the "fact stew" VanSledright (1995) describes.

After reading the trade book passages, the "friendly Columbus" element of Bill's positive conception remains stable as evidenced by Columbus's discovery of new lands and provision of new foods to Europe. Had I stopped the interview here, I might have concluded that Bill's response reflected a relatively thin historical interpretation, based on a weak content knowledge background, and a literal interpretation of one aspect from only one of the two readings.

What Bill seems to know and understand takes a dramatic turn, however, with my question about Columbus forcing Natives onto his ships. With that question, Bill's sense of differing perspectives on Columbus's actions and his identification with Columbus as a human being emerge. So, depending on where I look in the interview transcript, Bill's portrayal of Columbus varies as do the sources of his ideas. Based on his initial responses, Bill's content knowledge seems weak as he holds only a fragmented view of Columbus. Based instead on his initial interpretation, Bill appears to embrace a naively optimistic view of Columbus due largely to a literal reading of a portion of one of the texts. But based on his concluding interpretation, Bill seems to hold some sound insights into Columbus as a man, based on an empathic understanding of this historical actor as a human being.

As these points suggest, Bill's ideas about Columbus are neither static nor impermeable. He holds onto his initially positive view of Columbus through the readings until asked directly about Columbus's enslavement of Natives. He recalculates, concluding that Columbus was good *and* bad, a condition that he ascribes to himself and to humanity at large. The fluid nature of Bill's historical sense neither means that there is an orderly progression to his ideas, nor does it indicate that his ideas are in constant flux. Fluid, in this sense, means that Bill's ideas are susceptible to change and that big changes are just as possible as small ones.

The slippery nature of language. As noted earlier, there are a number of reasons for the instability of Bill's historical account. Most observers assume that blackening a bubble on a test score sheet does not well represent what students know. And yet, can we be any more certain based on what children say and what observers hear?

A general problem is the medium through which students express their ideas. Unlike the language of mathematics and the natural sciences, where words typically have specific referents, the language used in history and the social sciences is typically expressed in everyday terms. The

result: The language used in mathematics or physics, for example, is less susceptible to multiple interpretations than that used in history or geography.[ix] As Lowenthal (2000) notes, "History has no technical jargon and requires no grounding in some arcane aspect of nature or human nature" (p. 63). The very basis of historical representations then—language—contributes to the representation dilemma.

We can see the problematic nature of language as a means of representing historical ideas in Bill's interview as he characterizes Columbus (and himself) with words like "good," "bad," and "mean." Textbook apologists aside, any number of historians have arrived at a conclusion similar to Bill's—Columbus was "half good, half bad."[x] Although historians might smile and nod knowingly, do their reactions mean that Bill shares their construal? Terms like "good" and "bad" give broad interpretive cover: If Samuel Eliot Morrison chose to agree with this third-grader, should we conclude that they share the same knowledge and understandings? Bill's factual confusions suggest that he holds nothing like Morrison's evidence for a mixed assessment of Columbus. So, when Bill offers the more nuanced portion of his interpretation, what knowledge is he representing? His conclusion *sounds* like that reached by historians. But what if it is based more on his lived experience or what Levstik and Barton (1997) call "human sense," rather than on any fact-based analysis? If so, does this count as *knowing* history?

If one piece of the language dimension of the representation dilemma is the general nature of language, a second element hinges on the social and interpretative nature of representing ideas (Austin, 1975; Searle, 1989). Bernbaum (1962) states the problem generally: "The meaning of words varies with the context in which they are used, and also with the contexts in which speaker and listener have previously experienced the words" (p. 39). More specifically, Wertsch (2000) talks about the public nature of knowing history and concludes that "representing the past is dialogic rather than a monologic production by an independent agent" (p. 40).

Bill has some ideas about Columbus, but those ideas do not take shape until he and I enter into dialogue, and by doing so, two conditions impact Bill's representations. First, Bill's ideas *become* representations when he utters them and when I try to interpret them. Held silently or broadcast to the wind, Bill's ideas fail to take form as *representations* of his historical thinking. Once he offers his ideas publicly, my role as listener/interpreter begins. I listen and try to make sense of Bill's ideas; I ask a range of questions to probe his thinking, and I modify my initial interpretations continuously. Neither wholly Bill's nor mine, the emergent representations owe more to the conversational interaction between us than to an objective sense of the words themselves.

If the first condition of the social and interpretative notion of representations is that ideas are made in public and in interaction, a second condition is that those representations can change as the interaction continues. Had I asked but a single question about Columbus, Bill's response and my interpretation of that response would satisfy the demands of a historical representation. By continuing the interview with more questions, more responses, and more interpretations of those responses, it soon becomes clear that Bill's representation of Columbus is malleable. As noted earlier, the notion that the kinds of questions asked can influence the responses made (Dickinson & Lee, 1984; Nuthall & Alton-Lee, 1995; VanSledright, Kelly, & Meuwissen, 2006) raises questions about the certainty any listener/interpreter can feel about knowing what a speaker/writer knows.[xi]

Clearly, I have only scratched the surface of the language issues inherent in the representation dilemma.[xii] These issues loom large, but as we will see in the next section, they also loom differently for researchers than they do for teachers and for policymakers.

The need to evaluate students' representations. The fluid nature of students' representations and the slippery nature of language could be dismissed as academic musings if not for the evaluative role teachers and policymakers play. The third piece of the representation dilemma, then, is the practical need for teachers and policymakers to judge the historical representations students offer.

As a researcher, my uncertainty about what Bill knows arises in the context of having far more information about what he knows than is typical. I spent 25 minutes *trying* to understand how Bill makes sense of Christopher Columbus. By contrast, teachers typically rely on responses to a few questions during a classroom recitation and/or on an end-of-unit test. And policymakers may try to make sense of a child's knowledge based on a response to a single multiple-choice question. Bill faces no ramifications if I misinterpret what he knows; he faces far more serious consequences if his teacher and the state test maker get it wrong.

The Testing Dilemma

From one vantage, the need to evaluate students' historical representations might be viewed as part of the *testing dilemma*. But testing is only one way to assess students' knowledge and understanding (Grant, 2006a). There is no perfect assessment; the problems of the representation dilemma do not diminish, if, for example, one moves from multiple-choice tests to portfolios. The rise of high-stakes history testing, however, adds a new dimension to the issue of assessment, and, I argue, that it represents a new dilemma for communities of researchers, teachers, and policymakers.

Taking the long view, interpreting what Bill knows and understands is intriguing, especially when viewed under the presumptions that guide each community of practice. This generous view suffers, however, as the actors in each community now work within a new context: high-stakes history testing. In this context, I envision five problems that combine to create a testing dilemma, a dilemma that both incorporates and pushes beyond traditional psychometric worries of validity and reliability.

Elements of the Testing Dilemma

Before exploring the elements of the testing dilemma, let me offer some context. High-stakes testing of history content is not new. The Regents testing program in New York, for example, was implemented over 100 years ago. Recent revisions to the program have not intruded on the central premise that, in order to graduate from high school, students must pass several Regents exams.[xiii] As a gatekeeper to graduation, the Regents testing program is an example of a high-stakes testing context. Currently, close to 50% of the states administer standards-based social studies tests of one type or another (Grant, 2006b).[xiv] Of those states, ten have coupled tests with a high-stakes consequence for students and/or their teachers. The source of much new testing and much higher stakes generally, the No Child Left Behind legislation waffles on history testing (Gaudelli, 2006; Grant, 2006b) and so has created a bit of uncertainty among state policymakers. Maine policymakers, for example, recently pulled back their state

social studies exam in deference to the NCLB emphasis on reading and mathematics. At the same time, however, Tennessee and West Virginia policymakers added new social studies tests to their assessment programs (Grant, 2006b).

Although the future of social studies assessments is unpredictable, it seems reasonable to conclude that some sort of high-stakes testing will continue (Gaudelli, 2006). As it does, communities of teachers, researchers, and policymakers have much to consider as the problematic elements of the testing dilemma surface.

One problem is the *authenticity* of state-level social studies exams. In many ways, this problem represents the general issue of construct validity (Messick, 1988). As Horn (2006) notes, "In the case of history tests, assessing construct validity determines the extent to which a test accurately and adequately captures how much a student knows or can do related to the specific areas of interest such as U.S. or world history being measured" (p. 64). The coherence between a test and what students know is a key issue, one that both psychometricians and social studies educators are right to worry about (Grant, Gradwell, & Cimbricz, 2004). But in the questions that Horn (2006) poses as central to determining the construct validity of a test is one that has been largely ignored: "How is a competent student of history defined?" (p. 65). Up to this point in time, the answer to this question has defaulted to whatever cut score psychometricians have determined. Many observers rail against what seem like arbitrary pass/fail distinctions, but the fact that teachers, researchers, and policymakers have yet to articulate what a "competent student of history" is leaves psychometricians with little to go on. State-level test developers are rightly open to criticism that their efforts offer only a glimpse into what students know and understand. Their efforts are only as good, however, as the constructs available to them. And on that score, social studies teachers, researchers, and policymakers have not been of much assistance.

Application of test scores is a second aspect of the testing dilemma. The application problem involves the all-too-common occurrence of children who know some facts and can make some evidence-based inferences, but who may see no application of history or history testing to their lives.

In some ways, this problem is the flip side of the authenticity problem. That problem—defining the "competent student of history"—represents the adult point of view, i.e., how teachers, researchers, and policymakers define student historical competence. The application problem asks a similar question but from the student perspective. In other words, what happens if students see no validity to the tests they are taking? Social studies educators have long fought to help students perceive the relevance of history to their daily lives. If students come to associate history with a handful of multiple-choice questions, not only may they reject the importance of the exam, but they may reject the possibilities for history writ large.

British researcher Denis Shemilt (2000) argues that educators ask students to do "reverse somersaults" when teachers tell them, on the one hand, that there is "no single right answer to any of the really significant questions in history and that pupils must work things out for themselves" (p. 98), but then follow that invitation with a host of conditions that result in a sense that not all answers are equally valued. In the US, educators further complicate this reverse somersault with standards documents asserting the idea that no single historical account of an event can be constructed while, at the same time, administering tests which appear to privilege a series of single right answers.

A third element of the testing dilemma concerns the *scale* on which student responses are evaluated. State-level policymakers can report the results of students' exam performance in a

number of ways—a simple pass/fail mark based on a singlecut score, a fail/pass/mastery distinction based on two cut scores, or a rubric-like grade based on a scale with 3-5 cut scores. Making finer and finer distinctions in students' test performance *seems* as though it ought to help define what a "competent student of history" knows and understands, but does it? Is the competent student one who barely passes the New York Regents exam in U.S. history, for example, or is that designation reserved only for those who score at the mastery level?[xv]

Lee and Shemilt (Lee, 2005; Lee & Shemilt, 2004; Shemilt, 2000) are attempting to describe and label the "progression" of students' historical sense-making. Their work promises to offer insights useful to all communities of practice, but Lee and Shemilt (2004) warn that their ideas cannot be directly translated into standardized tests for two reasons: Most mandatory tests assume an "all or nothing" stance, and they fail to "consider how pupils make sense of what they learn and are asked to do" (p. 17). Lee and Shemilt conclude that "research-based models of progression are neither as comprehensive nor as utilitarian" as standardized assessments typically are expected to be.[xvi] So, test developers are handicapped on two dimensions—no agreed-upon definition of students' competence in history and no agreed-upon scale by which students' knowledge and understandings might be assessed.

Underlying these first three elements of the testing dilemmas is a number of factors, but perhaps none is as key as the problem of *development*. Piaget's (1962) stages of children's cognitive development have been enormously influential in the field of education. Those stages describe children's thinking as developing sequentially from concrete-to-abstract, from simple-to-complex, and from known-to-unknown (Lyle, 2000). In history, however, researchers challenge the notion that thinking develops in age-dependent and hierarchical stages (Booth, 1980; Dickinson & Lee, 1984; Levstik, 1986; Seixas, 1994, 1997; Shemilt, 1980). As Dickinson and Lee (1984) note, a stage view of students' historical thinking is problematic in a number of ways:

> The age-stage relation is only statistical, and, more important, the stages themselves are necessarily simplified models. Even where a child's thinking may in general be allocated to a given stage, his performance on any particular task may fluctuate widely according to the nature of the task, the variety of his experience, and the surrounding circumstances. (p. 118)

Lee and Shemilt (2004) find no evidence that historical knowing and understanding follows a "sequence of ladder-like rungs that every student must step on as he or she climbs" (p. 16).[xvii] Elements of students' historical thinking, such as their use of evidence, can be identified and outlined through a set of assessment markers. For example, students who see artifacts, in a "pictures of the past" sense, "treat potential evidence as if it offers direct access to the past" (p. 21). By contrast, students who demonstrate an "evidence in context" stance understand that "we must know what a source meant to those by and for whom it was produced" (p. 21). The value of models like the one Lee and Shemilt offers, then, emerges as a means of predicting the kind and range of ideas that teachers, researchers, and policymakers may encounter. That said, if, when, and how students move from a "pictures of the past" position to one of "evidence in context," Lee and Shemilt argue, can be identified in the aggregate, but not for individual students—a major problem for test developers. Moreover, models of progression that highlight a single element—e.g., students' use of evidence—may excite researchers, but offer only limited insight

into the bigger issues that teachers and policymakers wrestle with in the context of state-level exams.

One last dimension of the testing dilemma is *audience*. Children's ideas about history may or may not hold value when applied to their own lives. In the testing context, however, children's ideas have value only if they are perceived as such by the adults who interpret their responses. But what happens if those adults do not agree on or are uncertain about what constitutes historical knowing and understanding?

This last question can no longer be ignored. Before the advent of high-stakes testing, communities of teachers, researchers, and policymakers could disagree about the meaning of children's historical knowledge without serious implications for students Grant, 2003; Shulman, 1983). High-stakes testing ratchets up the consequences for such disagreements, however, as students and their teachers face sanctions for poor performance Grant, 2006a). In this context, the question "How do *we*—teachers, researchers, and policymakers—know what children know about history?" becomes critical, for how can we—the members of teacher, researcher, and policymaker communities—make judgments about children's school fates without a clear sense of what counts as knowing?

So I conclude with a challenge to the constituent communities of practice to come together and to (a) talk through the issues and conditions at hand, (b) discuss a set of relevant markers of historical knowing and thinking, (c) select a set of those markers that could be reasonably assessed in low and high-stakes settings and for students of mixed age and ability, and (d) create a set of criteria by which students' performance on those marker assessments can be judged. Any of three professional organizations—the National Council for the Social Studies (NCSS—teachers), the College and University Faculty Assembly (CUFA—researchers), or the Council of State Social Studies Supervisors (CS4—policymakers)—might lead this effort. But this conversation can be put off no longer.

References

Adler, D. (1991). *A picture book of Christopher Columbus*. New York: Holiday House.

Austin, J.L. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.

Baker, E. (1994). Learning-based assessments of history understanding. *Educational Psychologist, 29*(2), 97-106.

Barton, K., & Levstik, L. (1998). "It wasn't a good part of history": National identity and students' explanations of historical significance. *Teachers College Record, 99*(3), 478-513.

Bernbaum, G. (1962). Language and history teaching. In W.H. Burston & C.W. Green (Eds.), *Handbook for history teachers* (pp. 38-46). London: Metheun.

Booth, M. (1980). A modern world history course and the thinking of adolescent pupils. *Educational Review, 32*, 245-257.

Dickinson, A.K., & Lee, P.J. (1984). Making sense of history. In A. Dickinson, P. Lee, & P. Rogers (Eds.), *Learning history* (pp. 117-153). London: Heinemann.

Fish, S. (1980). *Is there a text in this class?* Cambridge, MA: Harvard University Press.

Grant, S. G. (2006a). *Measuring history: Cases of high-stakes testing across the US.*. Greenwich, CT: Information Age Publishing.

Grant, S. G. (2006b). The state of state-level history testing. In S.G. Grant (Ed.), *Measuring history: Cases of high-stakes testing across the US* (pp. 9-27). Greenwich, CT: Information Age Publishing.

Grant, S. G. (2003). *History lessons: Teaching, learning, and testing in U.S. high school classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.

Grant, S. G. (1995). Teaching history: Curricular views from California and the United Kingdom. *Journal of Social Studies Research, 19*(2), 17-27.

Grant, S. G., Gradwell, J. M., & Cimbricz, S. K. (2004). A question of authenticity: Examining the document based question on the New York state global history and geography Regents exam. *Journal of Curriculum and Supervision, 19*(4), 309-337.

Gaudelli, W. (2006). The future of high-stakes history assessment: Possible scenarios, potential outcomes. In S.G. Grant (Ed.), *Measuring history: Cases of high-stakes testing across the US* (pp. 321-334). Greenwich, CT: Information Age Publishing.

Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure. Th*eory Into Practice, 42*(1), 30-41.

Horn, C. (2006). The technical realities of measuring history. In S.G. Grant (Ed.), *Measuring history: Cases of high-stakes testing across the US* (pp. 57-74). Greenwich, CT: Information Age Publishing.

Krensky, S. (1991). *Christopher Columbus*. New York: Random House.

Lave, J. (1991). Situating learning in communities of practice. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 63-82). Washington, DC: American Psychological Association.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Lee, P. (2005). Putting principles into practice: Understanding history. In M.S. Donovan & J.D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 31-78). Washington, DC: National Academies Press.

Lee, P., & Shemilt, D. (2004). A scaffold not a cage: Progression and progression models in history. *Teaching History, 113*, 13-24.

Levstik, L. (1986). The relationship between historical response and narrative in a sixth-grade class. *Theory and Research in Social Education, 24*(1), 1-19.

Levstik, L., & Barton, K. (1997). *Doing history: Investigating with children in elementary and middle schools*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lowenthal, D. (2000). Dilemmas and delights of learning history. In P. Stearns, P. Seixas & S. Wineburg (Eds.), *Knowing, teaching, and learning history: National and international perspectives* (pp. 63-82). New York: New York University Press.

Lyle, S. (2000). Narrative understanding: Developing a theoretical context for understanding how children make meaning in classroom settings. *Journal of Curriculum Studies, 32*(1), 45-63.

Messick, S. (1988). Assessment in the schools: Purposes and consequences. In P. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice* (pp. 107-125). Berkeley, CA: McCutchan.

Morison, S.E. (1955). *Christopher Columbus, mariner*. Boston: Little, Brown.

National Center for History in the Schools. (1994). *National standards for United States history*. Los Angeles: Author.

Nuthall, G., & Alton-Lee, A. (1995). Assessing classroom learning: How students use their knowledge and experience to answer classroom achievement test questions in science and social studies. *American Educational Research Journal, 32*(1), 185-223.

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. New York: W. W. Norton.

Rogers, V., & Stevenson, C. (1988). How do we know what kids are learning in school? *Educational Leadership, 45*, 68-75.

Searle, J.R. (1989). *Speech acts: An essay in the philosophy of language*. New York: Cambridge University Press.

Seixas, P. (1994). Students' understanding of historical significance. *Theory and Research in Social Education, 22*(3), 281-304.

Seixas, P. (1997). Mapping the terrain of historical significance. *Social Education, 61*(1), 22-27.

Shemilt, D. (1980). *Evaluation study: Schools council history 13-16 project*. Edinburgh: Holmes McDougall.

Shemilt, D. (2000). The caliph's coin: The currency of narrative frameworks in history teaching. In P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching, and learning history* (pp. 83-101). New York: New York University Press.

Shulman, L. (1983). Autonomy and obligation. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 484-504). New York: Longman.

VanSledright, B. (1995). "I don't remember--the ideas are all jumbled in my head" Eighth graders' reconstructions of colonial American history. *Journal of Curriculum and Supervision, 10*(4), 317-345.

VanSledright, B. (2002). *In search of America's past*. New York: Teachers College Press.

VanSledright, B., Kelly, T., & Meuwissen, K. (2006). Oh, the trouble we have seen: Researching historical thinking and understanding. In K. Barton (Ed.), *Research methods in social studies education* (pp. 207-233). Greenwich, CT: Information Age Publishing.

Wertsch, J. (2000). Is it possible to teach beliefs, as well as knowledge about history? In P. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching, and learning history* (pp. 38-50). New York: New York University Press.

Author Notes-

---

[i] Although in some contexts there are useful differences between "history" and "social studies," I will use these terms interchangeably in this paper.

[ii] Thanks to Bruce VanSledright for insights on this construct. Note, however, that apart from teachers, researchers, and policymakers, other relevant communities of practice may include historians, psychometricians, school administrators, parents, and the general public.

[iii] The community labels "teacher" and "researcher" are self-explanatory. I use the "policymaker" label, however, to represent those state-level officials who have direct or indirect responsibility for designing, administering, and/or interpreting the results of state-level social studies exams. I use the label "test developer" later in this paper to refer to those directly responsible for the creation of state tests.

[iv] I pulled this interview data from that of a larger study I have been doing on how children understand complex human behavior.

[v] I use the language of "historical knowledge and understanding" as a placeholder for the many types of historical thinking as outlined, for example, in the *National Standards for History* (National Center for History in the Schools, 1994).

[vi] The genesis of this project is an activity I do with the students in my elementary social studies methods class. Based on that activity, I initiated a small research study in which I have interviewed 30 K-5 children on their ideas about the actions of historical actors like Christopher Columbus.

[vii] Interestingly enough, fewer than half of the students interviewed in the larger study commented on the idea of forcing Natives to leave on their own after the first reading. Moreover, their views were mixed as to whether forcing Natives to leave their homes was a "bad" thing. Some students thought the captives "lucky" because they were able to see a new land, while others thought that being forced to do something was of no great concern: They are frequently "forced" to do things they do not want to do by various adults—parents, teachers, principals.

[viii] VanSledright (2002) offers a useful discussion of the latter idea, which he calls "historical contextualization."

[ix] Of course, this generalization holds only so far. Physicists, for example, use the everyday term "light," but have discovered that, depending on how it is measured, it may be a ray or a wave. By contrast, geographers have only one definition for the term, "contour line." See Bernbaum (1962) on the general problem of language use in history.

[x] An interesting example is Samuel Eliot Morison (1955). Immediate reaction to his classic text, *Christopher Columbus, Mariner* focused on the negative aspects that Morison laid out, namely the genocide perpetuated by Columbus's men. In later years, Morison was portrayed as a Columbus apologist for his steady regard for Columbus's nautical skills.

[xi] Policymakers are no less susceptible to problems in language use (Lee & Shemilt, 2004)

[xii] Another huge language issue involves non-native speakers of English. The representation dilemma evident in the context of Bill's responses can only be compounded with children who are learning English at the same time they are learning history.

[xiii][xiii] The list of exams students must take and pass include one test in English Language Arts, one in mathematics, one in science, and two in history (Global History and Geography and US History and Geography).

[xiv] Since 1998, the number of states with social studies exams has fluctuated from 17-24 (Grant & Horn, 2006)

[xv] The cut score for passing the New York Regents exams in global and US history is 65; the cut score for mastery is 85.

[xvi] Interestingly enough, Lee and Shemilt (2004) make this claim in the context of the assessments developed from the British National Curriculum for history, which are far more ambitious than the tests typically developed in the US (Grant, 1995).

[xvii] Lee and Shemilt (2004) add, however, that, although Piaget's stage theories have not borne fruit in the study of history learning, it would be a mistake to discount all of Piaget's ideas.