

“Can You Make ‘Historiography’ Sound More Friendly?”: Towards the Construction of a Reliable and Validated History Teaching Observation Instrument

Stephanie van Hover, David Hicks, and Stephen Cotton
University of Virginia, Virginia Tech, and University of Virginia

Research in classrooms where history is taught indicates that the primary concern there is with consuming and reproducing events and details mostly found in books, as though interpretive practices, be they engaged in by historians, teachers, or students, simply did not exist...The standard textbooks, combined with lectures delivered by teachers, are considered definitive. Tests measure the results. The obsession appears to be with the products of historical study, not with the practice of doing it.¹

Bruce VanSledright, 2002

THE OPENING QUOTATION captures the “typical” student experience in the history classroom—an experience that often includes learning basic static factual knowledge by passively listening to lectures, reading textbooks, filling in worksheets, and preparing for tests.² Such an approach to teaching history does not reflect best practice in history education; rather, research and practitioner-based literature on high-quality teaching (and learning of history) promotes attention to the disciplinary nature of history, the “doing of history.”³ As the papers in this issue highlight, the literature recognizes the fluid, contested, and constructed nature of the inferential discipline of history and advocates preparing teachers to teach students to think historically through engaging in historical inquiry and to develop the knowledge, skills, and habits of mind necessary for active, informed, deliberative citizenship.

While the field of history education elucidates a clear and ambitious vision of high-quality history instruction,⁴ a current challenge for history educators (including teacher educators, curriculum specialists, and school-based history and social science supervisors) becomes how to illuminate and capture this when observing classrooms to research history instruction or to provide useful discipline-specific feedback to preservice (and inservice) history teachers. As Pianta and Hamre observe, descriptive and qualitative approaches to observation provide “rich, descriptive information about teachers’ practices and students’ experiences,” but can be idiosyncratic.⁵ They note that in recent years, there has been a “renewed emphasis on developing standardized classroom observational measures with adequate reliability and validity.”⁶ This call for high-quality standardized classroom observational measures comes at a time of national scrutiny of current teacher evaluation systems that do “little to help teachers improve or to support personnel decision making” and as consensus grows “that evidence of teacher contributions to student learning should be part of teacher evaluation systems, along with evidence about the quality of teacher practices.”⁷ And while a number of valid and reliable observation tools exist,⁸ it is worth noting that none focus specifically on the teaching of secondary history.

This paper introduces the structure of the *Protocol for Assessing the Teaching of History* (PATH), an instrument that provides one lens through which to observe secondary history teaching in order to provide a means for structured and focused observation of history teaching and learning with the goal of improving instruction. We make no claims that PATH is *the* way of teaching and learning history; rather, PATH initiates the conversation about how to capture and explore the specific teaching behaviors that the research and practitioner literature has shown to contribute to high-quality history instruction. PATH is an attempt to, as Grossman terms it, engage in the difficult work of “‘decomposition’ of practice—breaking down complex practice into its constituent parts for the purposes of teaching and learning.”⁹ The authors recognize that debates continue over the nature and purpose of history and its place in the K-12 curriculum;¹⁰ however, as history educators, our work emerges from and specifically recognizes that 1) the current policy-making context, at both state and national levels, explicitly advocates the establishment and implementation of teacher performance standards and evaluation criteria that emphasize student academic progress and pay attention to organization, delivery of instruction, and student assessment; 2) teacher education programs are under increasing pressure to collect reliable and valid observation data on preservice teachers for the purposes of state and national accreditation and to demonstrate “value-added” information¹¹; and 3) there exists an

emerging body of literature that seeks to study, describe, and provide images of high-quality history instruction.¹²

PATH

PATH¹³ is modeled on and informed by the Classroom Assessment Scoring System-Secondary (CLASS-S) (with written permission from CLASS's authors), an instrument developed to assess classroom quality.¹⁴ CLASS focuses on the interactions between students and teachers as the primary mechanism for student learning.¹⁵ PATH uses the same structure and scoring/coding approach as CLASS-S. Before using the tool, coders for PATH are trained on each dimension of a rubric through an in-depth manual that describes the specific teaching behaviors that comprise each dimension. PATH, a high-inference tool, is scored on a 7-point scale; scores assigned are based on alignment with anchor descriptions at Low (1, 2), Middle, (3, 4, 5), and High (6, 7).

To develop disciplinary-specific dimensions, we first conducted an extensive review of the literature on the teaching of history. We searched for work that could help us identify *observable* teacher and student behaviors that contribute to student learning.¹⁶ We identified and operationalized those descriptions of purposeful history instruction in order to distill and describe observable disciplinary-specific classroom practices.

From this review of the literature, six separate dimensions emerged: 1) Lesson Components, 2) Comprehension, 3) Narrative, 4) Interpretation, 5) Sources, and 6) Historical Practices (see Figure 1).

While we have developed these separate dimensions to help illuminate disciplinary-specific teaching, we also recognize that, due to the complexity inherent in teaching the discipline of history, these dimensions are neither mutually exclusive nor all necessarily observable within a single class observation. Under each dimension, we generated indicators and behavioral markers—the specific instructional behaviors or interactions that trained observers look for. Finally, we developed detailed descriptions of what “high,” “middle,” and “low” look like in order to develop portraits of practice and calibrate for inter-rater reliability.

In order to strengthen both content and face validity, history educators (in the United States and in the United Kingdom) and measurement experts reviewed the dimensions and provided critical feedback and suggestions. This feedback was solicited through e-mails, panel discussions, focus group meetings, and informal interviews. At the same time, the authors watched hundreds of hours of videotaped secondary history instruction; at first, these videos were used to identify observable teacher and student behaviors specific to history.¹⁷ Then, the videos were used to help capture

PATH Dimension	Brief Definition
Lesson Components	Assesses the structure and flow of the history lesson—the attention to objectives, assessment, and appropriate instructional approaches. Also, assesses attention to an overarching concept or framing historical question. ¹⁸
Comprehension	Assesses whether students understand the framework, key concepts, and content of history and are able to express this knowledge in different ways. ¹⁹
Narrative	Assesses the structure and flow of a narrative and whether students understand chronology, context, cause and effect, and that narratives are constructed. Narrative is defined as: Any contemporary written or verbal account (could include texts, lectures, websites, or films). ²⁰
Interpretation	Assesses the level of attention to the fluid and contested nature of history, as well as consideration of (if appropriate to the lesson objectives) agency, significance, diverse viewpoints, and perspective recognition. ²¹
Sources	Assesses the selection, accessibility, purpose, and level of analysis of historical sources that are used in a classroom, and, whether there is opportunity for meaningful historical inquiry. ²²
Historical Practices	Assesses whether general instructional practices (writing, discussion, simulations) are implemented in a way that are authentic and appropriate for the discipline of history. ²³

Figure 1: Overview of Dimensions, from Stephanie van Hover, Stephen Cotton, and David Hicks, *Protocol for Assessing the Teaching of History (PATH)* (Charlottesville, VA: University of Virginia, 2012).

“high,” “middle,” and “low” examples of each behavioral marker. Finally, the videos were coded using the instrument in order to collect inter-rater reliability data; collection of reliability data is ongoing, but early analysis indicates that we are meeting our target, an inter-rater reliability coefficient of .80. Additionally, PATH was informally field tested on secondary history preservice teachers.

PATH Dimensions, Behavioral Indicators, and Exemplars of its Use

As noted earlier, PATH is comprised of six dimensions (see Figure 1) that are broken into behavioral indicators or specific observable teacher or student behaviors. In order to highlight the nature and structure of PATH,

we detail two key dimensions—*Interpretation* and *Sources*—as illustrative exemplars. *Interpretation* refers to teaching in ways that recognize history as the product of the work of historians who, looking to represent the past, construct historical accounts from records and relics from the past. This dimension focuses attention on two specific behavioral indicators. The first behavioral indicator, “analysis and interpretation,” focuses on whether teachers create opportunities for students to understand that there are competing or conflicting accounts of history and to acknowledge the role of historians in constructing accounts or explanations of the past. This indicator also includes attention to such second-order concepts as historiography, empathy, agency, and significance. For example, does the teacher create spaces for discussions of historical agency, and the fact that historical events and processes are the result of decisions made by people in the past, as well as how historians ascribe significance to an event, person, or artifact? The second behavioral indicator, “perspective recognition,” focuses on the importance of examining historical context and recognizing that the values, attitudes, beliefs, and intentions of people in the past were the product of the time in which they lived.

In a lesson on the Holocaust, for example, a seventh-grade United States history preservice teacher who scored “high” on Interpretation introduced the lesson by explaining that historians continue to debate the precise date that the Holocaust began. She noted that although she would begin the discussion around 1935, her central argument for the day was a thesis she borrowed from a historian—that the Holocaust was a systematic, lengthy process that began as Hitler rose to power. Throughout the lesson, she tied particular events back to different accounts from historians and to her overarching thesis. During the lesson, students asked her, “Well, why didn’t [the Jewish people in Berlin] just run away?” This preservice teacher took the time to remind her students of the historical context and to ask them to consider things that constrained the actions of individuals. In short, her lesson addressed the role of historians in constructing history, explored the role of agency and empathy, and included attention to perspective recognition and specific historical contexts.

The dimension of *Sources* assesses the selection, accessibility, purpose, and level of analysis of historical sources used in the classroom. This dimension includes attention to three behavioral indicators—“selection and accessibility,” “purpose,” and “opportunity for inquiry.” Observers appraise whether the teacher selects, models, and scaffolds the use of appropriate and relevant sources. “Purpose” focuses on fidelity of implementation—the idea that there are different ways sources can be used in the classroom and that teachers should have a clear purpose and *achieve* that purpose. For example, sources might be used to introduce or close

a lesson, to illustrate a point or perspective, or to teach analysis skills in terms of opening up sources to a layers of inference approach—whether these purposes are achieved depends on the lesson objectives, the topic, the sources, the structure of the activity, and the presence of appropriately tailored reading questions. Finally, this dimension assesses meaningful opportunities for inquiry, where specific historical questions or issues to investigate are generated, and where students evaluate evidence and develop conclusions and arguments.

Observations of a lesson on the Italian Renaissance offer a “low” example of the Sources dimension. An inservice teacher chose to ask students to analyze *The Last Supper*, by Leonardo Da Vinci, as a means to review the key elements of the Italian Renaissance and to segue into the Northern Renaissance. The teacher made this purpose clear, describing his objectives as he introduced the activity. He then asked students to perform a modified AP-PARTS (i.e., Author, Place and Time, Prior Knowledge, Audience, Reason, The Main Idea, Significance) on the painting. While this teacher selected an appropriate source and provided a scaffold, he did not model or explain the modified AP-PARTS. The activity fell apart when students became confused about the idea of “Author” and “Audience” when analyzing the painting. When he realized students were confused, the teacher instead began to lead a question-answer session about the painting. However, in this discussion, rather than focus on the artist, the colors used, or the style or form of the painting, the teacher told students the biblical story of the Last Supper. Thus, the activity—using a historical source (*The Last Supper*) as a way to illustrate an idea (humanism)—did not achieve its purpose due to a lack of modeling, ill-fitting analysis questions, and follow-up questions unrelated to the lesson objectives (lack of alignment).

For these lessons and others, we envision the purpose and appropriate use of PATH as helping to facilitate positive discussions between an expert/trained observer and a preservice (or inservice) teacher to support the improvement of pedagogical practices that connect theory and practice in the history classroom. PATH is not designed as a one-time observational tool, but as a lens through which to begin to systematically and clearly connect the dots between theory and practice in order to support professional growth *over time*. Thus, it can be used for both preservice teachers and inservice teachers as a means of ongoing assessment of teachers’ instructional practice. PATH moves beyond the type of generic observational tools that fail to illuminate the nature and quality of discipline-specific pedagogical practices used to support learning as preservice students move throughout a teacher education program.

Discussion and Significance

The quotation in the title of our paper—“Can you make ‘historiography’ sound more friendly?”—was a question posed to us by a measurement expert and reflects the challenges involved in creating a reliable and validated observation instrument in history education. This paper focuses on the difficult task of balancing the very different “worlds” of measurement construction and history education and making constructs like historiography, narrative, significance, agency, and empathy observable and “friendly for observation.” The dimensions, indicators, and behavioral markers require “pulling apart” the key underlying disciplinary dimensions and constructs that make up the teaching of the “doing of history.” As history educators, we find it difficult to think in discrete dimensions as we observe history teaching—however, as researchers interested in creating a reliable and valid tool, it is necessary.

Preparing teachers to teach the inferential discipline of history is a complex and dynamic endeavor. The complexity of trying to capture effective history teaching in different contexts is extremely difficult, and those who try to negotiate that terrain do so at their own risk. While we recognize that our initial efforts may well be questioned, the attempt in and of itself is important. In a national educational context focused on teacher evaluation and value-added measures, PATH offers one approach to providing specific feedback to teachers of secondary history and supporting teacher education. In the field of history education, no validated research-based observation instruments currently exist. This is a gap that needs to be filled. If history educators are not ready, willing, or able to begin to engage in such work, the danger exists that others who know nothing of the discipline of history will look to fill the gap by developing more generic and less disciplinary-aware observational instruments. Within the current context, the purpose of PATH is to offer one measure that focuses on what history teachers should be doing to support student learning and to inform the preparation of preservice teachers.

Notes

1. Bruce VanSledright, “Confronting History’s Interpretive Paradox While Teaching Fifth Graders to Investigate the Past,” *American Educational Research Journal* 39, no. 4 (Winter 2002): 1091.

2. Keith Barton, "History: From Learning Narratives to Thinking Historically," in *Contemporary Social Studies: An Essential Reader*, ed. William B. Russell III (Charlotte, NC: IAP, 2011), 119-138; Bruce VanSledright, *The Challenge of Rethinking History Education: On Practices, Theories, and Policy* (New York: Routledge, 2011).
3. See Rosalyn Ashby, Peter Lee, and Dennis Shemilt, "Putting Principles into Practice: Teaching and Planning," in *How Students Learn: History in the Classroom*, ed. M. Suzanne Donovan and John D. Bransford (Washington, D.C.: National Academies Press, 2005), 79-178; Robert B. Bain, "'They Thought the World Was Flat?': Applying the Principles of *How People Learn* in Teaching High School History," in *How Students Learn: History in the Classroom*, ed. M. Suzanne Donovan and John D. Bransford (Washington, D.C.: National Academies Press, 2005), 179-213; S. G. Grant, *History Lessons: Teaching, Learning, and Testing in U.S. High School Classrooms* (Mahwah, NJ: Lawrence Erlbaum Associates, 2003); VanSledright, "Confronting History's Interpretive Paradox."
4. Grant; VanSledright, "Confronting History's Interpretive Paradox"; Bruce VanSledright, "Fifth Graders Investigating History in the Classroom: Results From a Researcher-Practitioner Design Experiment," *Elementary Education Journal* 103, no. 2 (November 2002): 131-160; Bruce VanSledright, *In Search of America's Past: Learning to Read History in Elementary School* (New York: Teachers College Press, 2002); VanSledright, *The Challenge of Rethinking History Education*; Sam Wineburg, *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (Philadelphia, PA: Temple University Press, 2001).
5. Robert C. Pianta and Bridget K. Hamre, "Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity," *Educational Research* 38, no. 2 (March 2009): 110.
6. Ibid.
7. Linda Darling-Hammond, Audrey Amrein-Beardsley, Edward Haertel, and Jesse Rothstein, "Evaluating Teacher Evaluation," *Phi Delta Kappan* 93, no. 6 (March 2012): 8.
8. See Thomas Kane and Douglas Staiger (lead authors of the Measures of Effective Teaching Project), *Measures of Effective Teaching: Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains* (Seattle, WA: Bill & Melinda Gates Foundation, 2012).
9. Pamela Grossman, "Framework for Teaching Practice: A Brief History of an Idea," *Teachers College Record* 113, no. 12 (December 2011): 2839.
10. Keith Barton and Linda Levstik, *Teaching History for the Common Good* (Mahwah, NJ: Lawrence Erlbaum Associates, 2004); VanSledright, *In Search of America's Past*; Wineburg.
11. See, for example, Mary Kennedy, Soyeon Ahn, and Jinyoung Choi, "The Value Added by Teacher Education," in *Handbook of Research on Teacher Education: Enduring Issues in Changing Contexts*, ed. Marilyn Cochran-Smith, Sharon Feiman-Nemser, and D. John McIntyre, third ed. (Mahwah, NJ: Lawrence Erlbaum Associates, 2008) 1249-1273.
12. Barton and Levstik; Grant; S. G. Grant and Jill Gradwell, eds., *Teaching History with Big Ideas: Cases of Ambitious Teachers* (New York: Rowman & Littlefield Education, 2010); VanSledright, "Confronting History's Interpretive Paradox"; VanSledright, *The Challenge of Rethinking History Education*.
13. The development of PATH is funded by the "My History Partner" federal Teaching American History grant.
14. Robert Pianta, Bridget Hamre, and Susan Mintz, *Classroom Assessment Scoring System: Secondary Manual* (Charlottesville, VA: Teachstone, 2011).

15. Kane and Staiger, 19.

16. Literature reviewed includes research studies, theoretical pieces, the *National Standards in History* (USA), the *National Curriculum in History* (United Kingdom), the National Research Council of the National Academies' *How Students Learn History*, websites dedicated to the study/reflection on teaching history, and other benchmark and standards documents.

17. Archival videotapes are from the "My History Partner" Teaching American History grant. The grant is providing financial support for the initial development of this measure. The authors of PATH have been influenced by the ideas and approach of the Classroom Assessment Scoring System (CLASS). For more information on CLASS, go to <<http://www.teachstone.org/about-the-class/>>.

18. John D. Bransford, Ann L. Brown, and Rodney R. Cocking, eds., *How People Learn: Brain, Mind, Experience, and School* (Washington, D.C.: National Academies Press, 2000); Peter Lee, "Putting Principles into Practice: Understanding History," in *How Students Learn: History in the Classroom*, ed. M. Suzanne Donovan and John Bransford (Washington, D.C.: National Academies Press, 2005); Grant and Gradwell; Thomas Estes, Susan Mintz, and Mary Alice Gunter, *Instruction: A Models Approach* (Boston, MA: Allyn & Bacon, 2010); Pianta, Hamre, and Mintz; Grant Wiggins and Jay McTighe, *Understanding by Design* (Alexandria, VA: ASCD, 2005).

19. Barton; Pianta, Hamre, and Mintz; Grant; Lee.

20. Barton; Barton and Levstik; Christine Counsell, "Historical Knowledge and Historical Skills: A Distracting Dichotomy," in *Issues in History Teaching*, ed. James Arthur and Robert Phillips (London, U.K.: Routledge, 2000), 54-71; Bruce Larson and Tim Keiper, *Instructional Strategies for Middle and High School* (New York: Routledge, 2007); Lee; John Tosh with Sean Lang, *The Pursuit of History* (London, U.K.: Pearson Longman, 2006).

21. Bain; Barton; Sarah Brooks, "Historical Empathy as Perspective Recognition and Care in One Secondary Social Studies Classroom," *Theory and Research in Social Education* 39, no. 2 (Spring 2011): 166-202; O. L. Davis, Jr. and Elizabeth Yeager, eds., *Historical Empathy and Perspective Taking in the Social Studies* (Oxford, U.K.: Rowman & Littlefield Publishers, 2001); George Iggers, *Historiography in the 20th Century: From Scientific Objectivity to the Postmodern Challenge* (London, U.K.: Wesleyan University Press, 1997); Peter Lee and Rosalyn Ashby, "Progression in Historical Understanding Among Students Ages 7-14," in *Knowing, Teaching, and Learning History: National and International Perspectives*, ed. Peter N. Stearns, Peter Seixas, and Sam Wineburg (New York: New York University Press, 2000), 199-222; Peter Lee and Dennis Shemilt, "The Concept that Dares Not Speak Its Name: Should Empathy Come Out of the Closet?" *Teaching History*, no. 143 (June 2011): 39-49; Stéphane Lévesque, *Thinking Historically: Educating Students for the Twenty-First Century* (Toronto, Canada: University of Toronto Press, 2009); Peter Seixas, "Teacher Notes: Benchmarks of Historical Thinking: A Framework for Assessment in Canada," Manitoba Education, <www.edu.gov.mb.ca/k12/cur/socstud/foundation_gr8/tns/tn1.pdf>; VanSledright, *The Challenge of Rethinking History Education*; Robert Williams, *The Historian's Toolbox: A Student's Guide to the Theory and Craft of History* (Armonk, NY: M. E. Sharpe, 2007).

22. Ashby, Lee, and Shemilt; Keith Barton, "Teaching History: Primary Sources in History—Breaking Through the Myths," *Phi Delta Kappan* 86, no. 10 (June 2005): 745-753; Keith Barton, "Research on Students' Ideas About History," in *Handbook of Research in Social Studies Education*, ed. Linda Levstik and Cynthia Tyson (New York: Routledge, 2008), 239-258; Barton and Levstik; Department of Education (UK), "History: The Statutory National Curriculum for England for Key Stage 3 History," <

education.gov.uk/schools/teachingandlearning/curriculum/secondary/b00199545/history>; David Sylvester, "Change and Continuity in History Teaching 1900-1993," in *Teaching History*, ed. Hilary Bourdillon (London, U.K.: Routledge, 1994), 9-23; VanSledright, "Confronting History's Interpretive Paradox"; VanSledright, "Fifth Graders Investigating History in the Classroom"; VanSledright, *In Search of America's Past*; VanSledright, *The Challenge of Rethinking History Education*.

23. Diana Hess, "Discussion in Social Studies: Is It Worth the Trouble?" in *Social Studies Today: Research and Practice*, ed. Walter Parker (New York: Routledge, 2010), 205-214; Larson and Keiper; Chauncey Monte-Sano, "Qualities of Historical Writing Instruction: A Comparative Case Study of Two Teachers' Practices," *American Educational Research Journal* 45, no. 4 (December 2008): 1045-1079.