U.S. DEPARTMENT OF EDUCATION

Teaching American History Evaluation

Final Report

Teaching American History Evaluation
Final Report

U.S. Department of Education

Office of Planning, Evaluation and Policy Development

Policy and Program Studies Service

Prepared by:

Phyllis Weinstock
Fannie Tseng

Berkeley Policy Associates
Oakland, Calif.

ଔ

Daniel Humphrey
Marilyn Gillespie
Kaily Yee

SRI International
Menlo Park, Calif.

2011

U.S. Department of Education
Arne Duncan
Secretary

Office of Planning, Evaluation and Policy Development
Carmel Martin
Assistant Secretary

Policy and Program Studies Service
Stuart Kerachsky
Director

August 2011

This report is also available on the Department's website at
http://www.ed.gov/about/offices/list/opepd/ppss/reports.html.

On request, this publication is available in alternative formats, such as Braille, large print, or CD. For more information, please contact the Department's Alternate Format Center at 202-260-0852 or 202-260-0818.

Contents

Exhibits

Acknowledgments

Executive Summary

In 2001, Congress established the Teaching American History (TAH) program, which seeks to improve student achievement by improving teachers' knowledge, understanding, and appreciation of traditional American history as a separate subject within the core curriculum. Under this program, grants are awarded to local education agencies (LEAs), which are required to partner with one or more institutions of higher education, nonprofit organizations, libraries, or museums. Grant funds are used to design, implement, and demonstrate innovative, cohesive models of professional development. In addition, grantees have been required to conduct project-level evaluations and have been encouraged to provide evidence of gains in student achievement and teacher content knowledge.

The U.S. Department of Education ("the Department") has awarded TAH grants annually since 2001, building to a cumulative total of approximately 1,000 TAH grants worth over $900 million. Grantees have included school districts in all 50 states, the District of Columbia, and Puerto Rico.

The current TAH study, which began in 2007, focuses on the 2004, 2005, and 2006 grantee cohorts, a total of 375 grantees. This study, conducted by Berkeley Policy Associates and SRI International, addresses the following questions:

- Is it feasible to use states' student assessment data to conduct an analysis of TAH effects on student achievement?

- What is the quality of TAH grantee evaluations?

  o Are TAH evaluations of sufficient rigor to support a meta-analysis of TAH effects on student achievement or teacher knowledge?

  o What are major challenges that impede implementation of rigorous grantee evaluations?

  o What are promising practices in evaluation, especially in the development of new assessments of student achievement in American history?

- What are strengths of TAH grantees' program designs and implementation?

  o What are major challenges that impede program implementation?

In order to address these questions, the study incorporated the following components:

- **Study of Feasibility of Analyzing State Data.** For the feasibility study, researchers reviewed the availability of states' American history assessment data and investigated the statistical power and validity of two rigorous quasi-experimental designs for analysis of TAH student outcomes.

- **Review of Quality of Grantee Evaluations.** Researchers reviewed 94 final evaluation reports made available by grantees funded in 2004, documented their research designs, and considered whether the evaluations could support a meta-analysis of TAH effects. As part of case study research, researchers also reviewed the ongoing evaluation practices of

the 16 grantees (of the 2006 cohort) visited and identified both challenges and promising approaches to evaluation.

- **Case Studies.** Case studies of 16 TAH grantees (selected from among 124 grantees in the 2006 cohort by matching eight pairs of grantees with similar demographics and different outcomes) could not associate practices with outcomes but provided in-depth qualitative data on grantee practices. Site visitors examined how TAH projects incorporated, adapted or struggled to implement high-quality professional development practices as defined in the professional development literature.

Findings

Feasibility Study

The feasibility study found that it was not feasible to use state data to analyze the effects of the TAH program on student achievement. The feasibility research, conducted in 2008, found that 20 states administered statewide, standardized student assessments in American history. Of the 20 states, many had revised or were in the process of revising their tests and did not administer these assessments consistently every year. The research team identified nine states with multiyear assessment data potentially sufficient for TAH outcomes analysis. A review of topics and formats of the assessments in these nine states indicated that the assessments addressed a range of historical topics and historical analysis skills that corresponded to the goals of the TAH projects in those states, as stated in grant proposals.[1]

Researchers considered two quasi-experimental designs, a regression discontinuity design and an interrupted time series design with a comparison group, for measuring the effects of the TAH program on student achievement. Preliminary estimates of statistical power suggested that power would be sufficient for these analyses. However, state data were ultimately available from only five states, out of a total of 48 states that had received TAH grants across the three funding cohorts included in the study. The limitations of the data compromised the rigor of the analyses as well as generalizability of the findings. Therefore, the analyses of TAH effects were infeasible.

Quality of Grantee Evaluations

Few grantees, either among the case study sites (2006 grantees) or among the 2004 grantees reviewed for a possible meta-analysis, implemented rigorous evaluation designs.

TAH evaluations were not sufficiently rigorous to determine the impact of the TAH program on achievement. The screening of 94 final evaluation reports of 2004 grantees for possible inclusion in the meta-analysis revealed that the great majority of evaluation reports either did not analyze student achievement outcomes, lacked controlled designs or did not provide detailed information about the sample, design and statistical effects. Of those evaluations

---

[1] The feasibility study did not include a detailed study of alignment of state assessments with TAH projects; full copies of assessments were not available. The assessment review was limited to comparison of broad topics and skills areas covered by the assessments and the projects.

with quasi-experimental designs, most used a post-test-only comparison group design and lacked adequate controls for preprogram differences in teacher qualifications and student achievement.

The case study research identified obstacles encountered by grantees in conducting evaluations, in particular the difficulty of identifying appropriate, valid, and reliable outcome measures for the measurement of student achievement and teacher content knowledge. For assessment of students, some evaluators noted that state-administered standardized assessments—if available—were undergoing review and revision and were not well-aligned with the historical thinking skills and content emphasized by the grants.[2] Other challenges faced by case study grantees in conducting outcomes-focused evaluations included identifying comparison groups for quasi-experimental evaluation and obtaining teacher cooperation with data collection beyond the state assessments, especially among comparison group teachers.

Some TAH evaluators were in the process of developing project-based assessments, including tests of historical thinking skills, document-based questions (questions based on analysis of primary source documents), assessments of lesson plans and student assignments, and structured classroom observations. However, many program directors and evaluators have noted that the development of project-based assessments requires a level of time, knowledge, and technical expertise that is beyond the ability of individual programs to undertake. Without further support, grantee-level evaluators have been unable to take these assessments to the next level of refinement and validation.

Strengths and Challenges of TAH Design and Implementation

Case studies of 16 TAH grantees documented ways in which TAH projects aligned, or failed to align, with principles of high-quality professional development as identified in the research literature and by experts in the field. These findings cannot be generalized beyond these 16 grantees, but they do offer insights into strengths and challenges of the TAH grants.

Strengths of the grantees are described below:

TAH professional development generally balanced the delivery of content knowledge with strengthening of teachers' pedagogical skills. TAH projects achieved this balance by helping teachers translate new history knowledge and skills into improved historical thinking by their students. Historians imparted to teachers an understanding of history as a form of inquiry, modeling how they might teach their students to closely read, question, and interpret primary sources. Some grantees then used master teachers to model lessons and to work directly with other teachers on lesson plans and activities that incorporated historical knowledge, resources, and skills that they were gaining through the grant.

Strong TAH project directors were those with skills in project management and in the blending of history content with pedagogy. Project participants praised project leaders who ensured that professional development was designed and delivered in ways that were useful for

---

[2] This evaluation did not systematically study the alignment between the grant and state standards. Case study respondents reported that professional development content was designed to be aligned with state standards, but the projects gave particular emphasis to "deepening teachers' understanding and appreciation" of American history (a primary goal of the TAH program stated in Department guidelines) rather than to strictly and thoroughly matching project activities to state standards.

instruction. These project leaders coordinated and screened project partners, provided guidance for historians on teachers' needs, and combined content lectures with teacher activities on lesson planning that linked history content to state and district standards.

Partnerships gave teachers access to organizations rich in historical resources and expertise, and were flexible enough to adapt to the needs of the teachers they served. The number, types and level of involvement of partners varied across the study sites. Partnerships praised by teachers connected teachers not only with historians but also with local historic sites, history archives, and primary sources. At some sites, partners engaged teachers in original research. Teachers in turn used this research to create lessons that engaged students in historical thinking.

TAH projects created varied forms of teacher networks and teacher learning communities and some made use of teacher networks to disseminate content and strategies to non-participants. The case study sites engaged teacher participants in a variety of informal and formal collaborations or "teacher learning communities." Some sites required participants to deliver training sessions to nonparticipants in their schools or districts. Networking and dissemination activities helped amplify and sustain the benefits of the grants.

TAH sites received praise from participants for establishing clear goals for teachers, combined with ongoing feedback provided by experts. Most sites required participants to make a commitment to attend professional development events, but a few sites went beyond this to hold teachers accountable for developing specific products. In one site, participating teachers were asked to sign a Memorandum of Understanding that clearly outlined the project goals, project expectations and requirements that teachers were required to fulfill in order to receive in-service credits, graduate credits, and a teacher stipend.

Key *challenges* experienced by TAH grantees are described below:

Most TAH case study sites were not implemented schoolwide or districtwide, and most received uneven support from district and school leaders. Obtaining strong commitments from district and school leaders was challenging for some project directors, particularly those administering multidistrict projects. Strategies that were successful included the creation of cross-district advisory committees and the linkage of TAH activities to school or district priorities such as improving student performance in reading and writing. In those grants with strong district-level or school-level support, teacher participation rates were higher and teacher networks were more extensive.

Most grantees struggled to recruit teachers most in need of improvement. Recruitment of American history teachers able to make a commitment to TAH professional development presented ongoing challenges for the case study sites. Project staff reported that it was especially difficult to recruit newer teachers, struggling teachers, and teachers with less experience in teaching history. Grantees used a wide variety of strategies to recruit teachers, such as widening the pool of participants to encompass larger geographic areas, more districts and more grade levels, and offering incentives, such as long-distance field trips, that sometimes resulted in high per-participant costs. Among strategies that grant directors reported to be successful were conducting in-person outreach meetings at schools to recruit teachers directly, and offering different levels of commitment and options for participation that teachers could tailor to their schedules and needs.

Conclusions and Implications

The Teaching American History program has allowed for productive collaborations between the K–12 educational system and historians at universities, museums, and other key history-related organizations. Respondents at 16 case study sites consistently reported that history teachers, who generally are offered fewer professional development opportunities than teachers in other core subjects, have deepened their understanding of American history through the TAH grants. Overall, participants lauded the high quality of the professional development and reported that it had a positive impact on the quality of their teaching. Teachers reported that they have increased their use of primary sources in the classroom and developed improved lesson plans that have engaged students in historical inquiry.

Extant data available for rigorous analyses of TAH outcomes are limited. TAH effects on student achievement and teacher knowledge could not be estimated for this study. Grantee evaluations that were reviewed lacked rigorous designs, and could not support a meta-analysis to assess the impact of TAH on student achievement or teacher knowledge. However, many of the project-based assessments under development by grant evaluators show potential and could be adapted for more widespread use. Given the limitations of state assessments in American history, these project-developed measures are worthy of further exploration and support.

Case study research did not find associations between TAH practices and outcomes but found key areas in which TAH program practices aligned with principles of quality professional development. The case studies found grantees to be implementing promising professional development programs that built on multifaceted partnerships, balanced history content with pedagogy, and fostered teacher networks and learning communities. In addition, some grantees and their evaluators were developing promising approaches to teacher and student assessment in American history. However, the case studies also found that Teaching American History grants often lacked active support from district or school administrators and were not well integrated at the school level. Grantees struggled to recruit a diverse range of teachers, particularly less experienced history teachers and those most in need of support.

Overall, the findings of this evaluation suggest a need for increased guidance for TAH grantee evaluations, teacher recruitment, and integration of the grants into ongoing school or district priorities.

Chapter 1
Introduction

The Teaching American History (TAH) grant program, established by Congress in 2001, funds competitive grants to school districts or consortia of districts to provide teacher professional development that raises student achievement by improving teachers' knowledge, understanding, and appreciation of American history. Successful applicants receive three-year grants to partner with one or more institutions of higher education, nonprofit history or humanities organizations, libraries, or museums to design and deliver high-quality professional development. Over the past decade, the Department has awarded over 1,000 TAH grants worth more than $900 million to school districts in all 50 states, the District of Columbia, and Puerto Rico.

Interest in the effectiveness and outcomes of the grants has grown, and as a result, in 2003, the Department introduced into the TAH grant competition a competitive priority to conduct rigorous evaluations. In addition, the Department sponsored an implementation study of the program, conducted by SRI International and focusing on the 2001 and 2002 cohorts of grantees. In 2005, the Department contracted with Berkeley Policy Associates to study the challenges encountered in implementing evaluations of the TAH projects.

In response to the 2005 study, the Department conducted a number of actions to encourage and assist the implementation of rigorous evaluations, such as: including a competitive preference priority encouraging applicants to propose quasi-experimental evaluation designs; providing grantees with ongoing technical assistance from an evaluation contractor; including evaluation as a strand at project director and evaluator meetings that highlighted promising evaluation strategies; and increasing the points for the evaluation selection criterion in the notice inviting applications.[3] In fiscal year (FY) 2007, the program included an option for applicants to apply for a five-year grant with the goal of obtaining better evaluation data. In the most recent competition (FY 2010), the program conducted a two-tier review of applications with a second tier comprised of evaluators reading and scoring only the evaluation criterion.

The current study, which began in 2007, is conducted by Berkeley Policy Associates and SRI International. The study addresses the following questions:

- Is it feasible to use states' student assessment data to conduct an analysis of TAH effects on student achievement?

- What is the quality of TAH grantee evaluations?

    o Are TAH evaluations of sufficient rigor to support a meta-analysis of TAH effects on student achievement or teacher knowledge?

    o What are major challenges that impede implementation of rigorous grantee evaluations?

    o What are promising practices in evaluation, especially in the development of new assessments of student achievement in American history?

---

[3] In addition, since this study has been underway, *Government Performance and Results Act* (*GPRA*) indicators have been revised to focus on participation tracking and use of teacher content knowledge measures.

- What are strengths of TAH grantees' program designs and implementation?
  - What are major challenges that impede program implementation?

In order to address these questions, the study focused on the 2004, 2005, and 2006 cohorts of grantees and incorporated the following components:

- **Feasibility Analysis.** The feasibility study reviewed the availability of states' American history assessment data and investigated the statistical power and feasibility of conducting several rigorous quasi-experimental designs to analyze TAH student outcomes.

- **Review of Evaluations.** Researchers reviewed the final evaluation reports of grantees of the 2004 cohort and considered whether the evaluations could support a meta-analysis of TAH effects. As part of case study research, researchers also reviewed the ongoing evaluation practices of the 16 grantees of the 2006 cohort, and identified both challenges and promising approaches to evaluation.

- **Case Studies.** Case studies of 16 grantees were designed to provide in-depth qualitative data on grantee practices. This study was informed by prior research on the accomplishments and challenges of the TAH program. In particular, the challenges of evaluating the program had been previously documented. Below we summarize findings of this earlier research, followed by a description of the research methods of the current study.

Previous Research on the Teaching American History Program

Earlier national studies of the TAH program have analyzed program implementation and implementation of evaluations but have not analyzed program outcomes. From 2002 to 2005, SRI International conducted an evaluation of the TAH program that focused on the 2001 and 2002 grantee cohorts. The study addressed three broad groups of research study questions: (1) the types of activities TAH grantees implemented; (2) the content of the activities, including the specific subjects and areas of American history on which projects focused; and (3) the characteristics and qualifications of teachers who participated in the activities.

The study found that the TAH projects covered a wide range of historical content, methods, and thinking skills. Grants were awarded to districts with large numbers of low-performing, minority, and poor students, suggesting that resources were reaching the teachers with the greatest need for improvement in their history teaching skills. However, a closer look at the academic and professional backgrounds of the TAH teachers showed that, as a group, they were typically experienced teachers with an average of 14 years of experience and were far more likely to have majored or minored in history in college than the average social studies teacher. Furthermore, while TAH projects did incorporate many of the characteristics of research-based, high-quality professional development, they rarely employed follow-up activities such as classroom-based support and assessment. An exploratory study of teacher lesson plans and other products also uncovered a lack of strong historical analysis and interpretation.

Although the SRI evaluation did not assess the impact of TAH projects on student or teacher learning, the evaluation did analyze grantee evaluations of effectiveness and found they often lacked the rigor to truly measure a project's effectiveness accurately. Ninety-one percent of the

project directors, for example, relied on potentially biased teacher self-reports to assess professional development activities, and substantially fewer used other methods like analyzing work products (64 percent) or classroom observations (48 percent).

A 2005 study by Berkeley Policy Associates examined project-level evaluations of nine of the 2003 TAH grantees and identified some potential challenges to conducting rigorous evaluations including: (a) difficulty in recruiting and retaining teachers, which led to serious delays in project implementation, infeasibility of random assignment, attrition of control group members, and small sample sizes; (b) philosophical opposition to random assignment; (c) conflict between project and evaluation goals (for example, honoring a school's philosophy of promoting teacher collaboration versus preventing contamination of control and comparison groups); and (d) difficulty in collecting student assessment data or identifying assessments that were aligned with the project's content. BPA recommended that the Department better define its priorities for evaluations of the TAH grants and extend the grant cycle so that recipients could devote the first six months to a year to planning, design, and teacher recruitment. Targeted technical assistance was recommended in order to improve the evaluation components of the grants and increase the usefulness of these evaluation efforts.

Measuring Students' Knowledge of American History

The Teaching American History Program was initiated in Congress in response to reports of weaknesses in college students' knowledge of American history (Wintz 2009). The National Assessment of Educational Progress (NAEP) is the single American history assessment that is administered nationally. The NAEP tests a national sample of fourth-, eighth- and twelfth-graders, and included an American history test in 1986, 1994, 2001, and 2006. Weak performance on this assessment has been a cause for concern, although noticeable improvements among lower performing students were in evidence between 1994 and 2006 (Lee and Weiss 2007). In general, NAEP results have pointed to weakness in higher order historical thinking skills, as well as students' limited ability to recall basic facts.

The measurement of trends in students' performance in American history is complicated by differences of opinion regarding what students should learn and the infrequent or inconsistent administration of assessments. The field is faced with a multiplicity of state standards, frequent changes in standards, and the low priority given to social studies in general under *the Elementary and Secondary Education Act (ESEA)* accountability requirements. Many states do not administer statewide American history assessments, and other states have administered them inconsistently.

While TAH grantees have been urged to meet *GPRA* indicators that are based on results on state assessments, such assessments are not always available. Further, TAH programs often emphasize inquiry skills and historical themes that are not fully captured through those state tests that are available. This study has examined these issues as they relate both to grantee-level evaluations and the national evaluation and presents strategies for developing promising approaches to measuring student outcomes.

Study Methods

Feasibility Study

The first task addressed by the study was the investigation of options for analysis of student outcomes of the TAH program using state data. Because little was known at the outset about the availability, quality, and comparability of student American history assessment data, a feasibility study—including research on state history assessments—was conducted. Based on early discussions among study team members, the Department, and the Technical Work Group, it was determined that both a regression discontinuity design and an interrupted time series design warranted consideration for use in a possible state data analysis. The feasibility study therefore was designed to research the availability and quality of student assessment data and to compare the two major design options and determine whether the conditions necessary to implement these designs could be met. Ultimately, the feasibility study found that student assessment data were available from a limited number of states and the analyses of TAH effects on student achievement could not be conducted.

Review of Evaluations

Among the three grantee cohorts included in the study, only the 2004 grantees had produced final evaluation reports in time for review; these final reports were potentially the best source of outcomes data for use in a meta-analysis. Of the 122 grantees in this cohort, 94 final reports were available for review. A three-stage review process was used to describe the evaluations, document their designs and methods, and determine whether they met criteria for inclusion in a meta-analysis.

Another component of the evaluation review was the review of evaluations of 16 case study programs, all of the 2006 cohort. Although final evaluation reports had not been completed at the time of the site visits, site visitors reviewed evaluation reports from the earlier years of the grants when available, and interviewed evaluators and project staff regarding evaluation designs and challenges in implementing the evaluations.

Case Studies

The goals of the case study task, as specified in the original Statement of Work, were to identify: 1) grantee practices associated with gains in student achievement; and 2) grantee practices associated with gains in teachers' content knowledge. In order to address these goals, the research team selected case study grantees using the selection process presented in detail in Appendix A. All case study grantees were selected from the cohort of TAH grantees funded in 2006. Using student history assessment data obtained from five states, researchers calculated regression-adjusted differences in average pre- and post-TAH assessment scores for all of the TAH grantee schools within these states. Four grantees with significant gains in students' American history scores were matched to four grantees within their states that exhibited no gains during this time. Selection of the second set of eight case studies, focusing on teacher content knowledge, was based on review of teacher outcomes data presented by grantees in the 2008 Annual Performance Reports (APRs). Four grantees with well-supported evidence of gains in teachers' content knowledge were matched to four grantees in their states that did not produce evidence of gains.

Researchers designed structured site visit protocols to ensure that consistent data were collected across all of the sites. The protocols were designed to examine whether and in what ways the TAH projects implemented or adapted key elements of professional development practice as delineated in the literature. Topics explored in the protocols included: project leadership; planning and goals; professional development design and delivery; district and school support; teacher recruitment and participation; evaluation; and respondents' perceptions of project effectiveness and outcomes. Site visits could not be scheduled until fall 2009, after the grant period was officially over, although most sites had extension periods and continued to provide professional development activities through the fall.[4] Researchers visited each of the case study grantees for two to three days. Site visitors interviewed project directors, other key project staff, teachers, and partners; reviewed documents; and observed professional development events when possible. Upon their return, site visitors prepared site-level summaries synthesizing all data collected according to key topics and concepts in the literature on effective professional development. The summaries provided the basis for cross-site comparisons and analyses.

Ultimately, no patterns in practices were identified that could clearly distinguish "high performing" and "typically performing" sites. However, the case study analysis identified areas of practice in which the case study sites exhibited notable strength and areas in which they struggled.

Content of This Report

In Chapters 2 through 5, the report presents findings of each of the study components. Chapter 2 presents results of the feasibility study. Chapter 3 presents findings of the review of grantee evaluations. Chapter 4 presents findings of the case study research on grantee practices. Finally, Chapter 5 presents conclusions and implications.

---

[4] Nine of the 16 case study grantees had received a no-cost extension to continue work beyond the grant period. Some entities had also received new grant awards. However, we focused on the activities of the 2006 grants in our interviews and observations.

Chapter 2
Feasibility of State Data Analysis to Measure TAH Effects

The evaluation team conducted feasibility research in order to identify options for the use of state assessment data to analyze the effects of the Teaching American History program on student achievement. The feasibility study was designed to: (1) determine the availability of state American history assessment data in the years needed for analysis (2005–08); and (2) identify the best analytic designs to employ should sufficient data be available. Although assessment data were available from five states, and two rigorous designs were considered, the data ultimately were insufficient to support analyses of the effect of TAH on student achievement.

State Assessments in American History

The feasibility analysis included research on states' administration of standardized assessments in American history, in order to identify those states from which student outcomes data might be available in the years needed for the TAH evaluation. Through a combination of published sources, Web research, and brief interviews, the study team addressed the following questions about student assessments in American history administered by states:

- At what grade levels are students assessed statewide in American history?

- Is it mandatory for districts statewide to administer the assessment?

- What are the major topic and skill areas covered by the test?

- Is the test aligned with state standards?

- Has the same (or equated) test been in place for at least three years?

- Is a technical report on the test available?

- If American history is only one strand of a larger social studies exam, is the American history substrand score easily identifiable?

Based on the information gathered, the study team identified nine states that were the most likely to have student assessment data that would meet the needs of an analysis of Teaching American History grant outcomes. These states administered statewide assessments in American history at either the middle school or high school level, or both, and had been administering comparable assessments for at least three years. Eleven other states had developed American history assessments, but the assessments were undergoing revision, were administered in only one or two of the years needed for analysis, or included American history as part of a broader assessment in history or social studies with no separate subscores in American history. Only nine states had consistent American history test score data in the years needed for analysis. A review of the topics, skills, and grade levels included in the tests determined that they broadly corresponded to the TAH project goals in those states.[5]

---

[5] The TAH grantees, in the cohorts under study, were not required to align their projects with state standards or to use state assessments to measure progress but were encouraged to do so if possible.

Given the data available, two analytic designs were considered: regression discontinuity design (RD) and interrupted time series design (ITS). RD is a more rigorous design than ITS, while ITS provides greater flexibility, greater statistical power, and would enable more precise targeting to participating schools.

Regression Discontinuity Design

The regression discontinuity design (RD) is considered the most internally valid non-experimental evaluation design available to evaluation researchers. However, the conditions under which RD can be applied are limited. A major factor in considering an RD study was that TAH grants are awarded using a well-documented point system with a consistent cutoff value. The RD design relies on inferences found in the neighborhood of the application cutoff point. Moreover, selection into the program group mimics the selection process in a random experiment, therefore yielding estimates that are free of any bias.

Through discussion with the TAH program office, it was established that TAH funding decisions are determined through a well-understood and well-documented independent selection process, in which a separately established cutoff point (score) is consistently used to distinguish applicants that are offered funding from those that are not. Therefore, funding decisions were made strictly according to the independent application scoring process, and there were no confounding factors with the funding assignment. The application score is a continuous variable with a sufficient range (for example, in 2004, scores ranged from 29.64 to 103.40) and a sufficient number of unique values. The TAH program office was able to provide documentation of the scoring system and also provided rank order lists of funded and unfunded applicants that made it possible to undertake preliminary calculations of statistical power of an RD design in the states with American history assessments. These preliminary calculations established sufficient confidence that an RD design warranted consideration.

However, because the TAH application scoring and funding process occurs on a yearly basis, it would be necessary to conduct an RD analysis separately for each of the three cohort years under study. This would limit the sample power for the analyses. Measuring an unbiased program effect using the RD technique relies upon the correct specification of the relationship between the application score (assignment variable), program participation, and the outcome. This relationship might differ across grant competitions in different years.

Another consideration for an RD analysis was that it would need to be based on districtwide student outcomes at grade levels relevant for the assessments. It would not be possible to drop from the analysis schools that did not have any participating teachers in the TAH program, because identifying schools with participating teachers was possible only for the funded applicants.[6] The review of applications conducted as part of the feasibility study had concluded that most applicants planned districtwide dissemination strategies, regardless of the proportion of history teachers committed to full participation. Researchers' estimate of school participation

---

[6] To the extent that unfunded district schools with teachers who would have participated in the TAH grant program differ from others, dropping schools for the funded districts while not dropping schools for the unfunded districts could bias the measured program effects. (The bias would be upward if higher-skilled and more motivated teachers are more likely to participate in the program and downward if lower-skilled and less motivated teachers are more likely to participate in the program.) Comparing districtwide outcomes in both funded and unfunded applicants gives us a fair test of the intervention as long as the power is sufficient to detect small ("diluted") effects.

rates in grantee districts in New York state, based on school lists provided by 21 grantees, found a 77 percent school participation rate across grantee districts.

Interrupted Time Series Design

The second analytic design under consideration was an interrupted time series design (ITS). This design uses the pattern of average scores prior to program implementation as the counterfactual with which the post-program test score patterns are compared. The main strength of the ITS model is its flexibility: it can be used to analyze data at many levels (state-, district- or school-levels, for example), and the design does not require large sample sizes to ensure adequate statistical power for the model. The model can be estimated with as few as three years of data, and it can be estimated on repeated cross-sections of data, instead of student-level panel data (student-level data that is linked by student across years). This aspect of the ITS model makes it especially well-suited for evaluating TAH program outcomes; because American history assessments are usually administered one time in high school, panel data on American history assessment outcomes generally do not exist. Another advantage of the ITS model was that it would be possible to target the analysis to participating schools.

However, the ITS model also has a number of weaknesses: the key weakness is lack of rigor. The ITS model has various threats to validity, including regression to the mean, selection bias, and history. Regression to the mean is a statistical phenomenon with multiperiod data whereby any variation from the mean in one period will tend to be followed by a result that is closer to the mean. If, for example, the year of TAH program implementation followed a year of lower-than-average results, any improvement in assessment results in the TAH program implementation that occurred because of the regression to the mean could be mistaken for a positive program effect. Selection bias would occur if participation in the grant program was correlated with unobserved characteristics that also affected American history assessment outcomes. If, for example, only higher-skilled teachers applied to receive grant program training, the ITS results could be biased. "History" threats could occur if implementation of the program coincided with another event or program that affects American history assessment outcomes. The proposed design included efforts to minimize these threats; for example, a comparison group was included in order to control for history threats. However, there is no way to entirely eliminate the threats to validity for the ITS model, and there exists the chance that the program impacts estimated with the ITS model would be influenced by other factors and could not be fully attributed to the program.

Challenges

Ultimately, it was determined that data were not sufficient to analyze the effects of TAH on student achievement. Primary considerations in this determination were:

- Because only a small portion of TAH grantees in the three cohorts would be represented in the data (assessment data were ultimately available from five states, out of 48 states from which applications were submitted over the three years), results of the analyses would not be generalizable to TAH projects in those cohorts overall.

- The limited proportion of grantees in the data also potentially compromised the ability, within the RD design, to model the correct relationship between TAH outcomes and receipt of the grant (applicant scores).

- The RD design necessarily would be conducted at the district level, and would measure a "diluted" effect of TAH; statistical power might not be sufficient to measure a very small effect.

- The ITS design, despite having more statistical power than the RD design, is less rigorous and could not be used to establish a causal relationship between the TAH program and patterns in students' American history test scores.

Chapter 3
Quality of TAH Grantee Evaluations

Since the inception of the TAH program, the Department has had a strong interest in determining the impact of the program on student and teacher learning. The primary available vehicle for assessing these outcomes has been the project evaluation each grantee must propose. Since 2003, through invitational priorities in the application process, the Department has urged grantees to conduct rigorous evaluations of their individual TAH programs using experimental and quasi-experimental evaluation designs that assess the impact of the projects on teacher knowledge and student achievement. The quality of the proposed evaluation is worth 25 out of the 100 points that may be received during the grantee selection process. This study addressed the following questions regarding TAH evaluations:

- What is the quality of TAH grantee evaluations?

    - Are TAH evaluations of sufficient rigor to support a meta-analysis of TAH effects on student achievement or teacher knowledge?

    - What are major challenges that impede implementation of rigorous grantee evaluations?

    - What are promising practices in evaluation, especially in the development of new assessments of student achievement in American history?

To address these questions, the study team conducted a thorough review of the final evaluation reports of the 2004 cohort of TAH grantees. In addition, the study team reviewed the ongoing evaluations of the 16 case study grantees, all members of the 2006 cohort of grantees, and interviewed case study project directors and evaluators to obtain in-depth information on evaluation methods and challenges encountered.

Findings of this study suggest that few grantees in either cohort of grantees implemented rigorous evaluations. This chapter discusses challenges of conducting these evaluations as well as promising assessment approaches of the case study grantees. Key findings include:

- Most grantee evaluation reports that were reviewed used designs that lacked adequate controls and did not thoroughly document their methods and measures. The quality of TAH grantee evaluations was insufficient to support a meta-analysis.

- Challenges in conducting evaluations included difficulties in identifying appropriate, valid, and reliable outcome measures; identifying appropriate comparison groups; and obtaining teacher cooperation with data collection, especially among comparison group teachers.

- Some evaluators interviewed during the cases studies had developed project-aligned assessments of teaching and learning in American history that show promise and warrant further testing and consideration for wider dissemination.

Review of Grantee Evaluations

This section presents the results of a review of grantee evaluations conducted to describe the evaluation methods, assess their quality, and determine whether a meta-analysis was feasible. Evaluations from the 2004 cohort of grantees were the most recent local evaluations that were available and potentially represented the best sources of outcome information.

A meta-analysis combines the results of several studies that address a set of related research hypotheses. This is normally accomplished by identification of a common measure of effect size, which is modeled using a form of meta-regression. Meta-effect sizes are overall averages after controlling for study characteristics and are more powerful estimates of the true effect size than those derived in a single study under a given single set of assumptions and conditions. In order to qualify for inclusion in a meta-analysis, studies must meet standards for design rigor and must include the information needed to aggregate effect sizes.

The review process began with 94 final TAH Annual Performance Reports[7] that varied widely in the amount of detail they provided about their project goals, professional development experiences, classroom instructional practices, and student and teacher learning outcomes. Most of the reports provided limited detail about the features and implementation of the TAH projects. The researchers were able to identify a small number of reports that provided adequate detail about student achievement outcomes and features of the study design.

A three-stage screening process was used to determine whether evaluation reports qualified for a meta-analysis. During Stage 1 of the screening process, all TAH final Annual Performance Reports submitted by grantees were reviewed and described. For each, the researchers recorded a description of the research design and measures used and reached an overall judgment of study rigor based on the following criteria:

- Presence of American history student test score data.

- Use of a quasi-experimental or experimental research design.

- Inclusion of quantitative information that could be aggregated in a meta-analysis (i.e., sample size, means and standard deviations for student test scores for each group reported).

See Exhibit 4 (Appendix B) for the results of the Stage 1 screening, including the initial judgment of rigor for each of the 94 TAH projects. Of the 94 reports reviewed in Stage 1, 32 met the above criteria for rigor and were selected for the second stage of the screening.

The 94 evaluations also were screened to identify those that measured improvements in teacher content knowledge in American history. Unfortunately, only eight evaluations met minimal requirements for inclusion in a meta-analysis. Upon further examination of the eight candidates for inclusion, the measurement instruments employed were not validated, and none provided enough detail to allow for a meta-analysis. Overall, the vast majority of evaluations of teacher outcomes were limited to self-report.

---

[7] In some cases, final evaluation reports were produced separately as attachments to the final Annual Performance Reports; in most cases, all evaluation information was incorporated into the Annual Performance Reports.

In Stage 2 of the screening process, the 32 reports that met the Stage 1 criteria were the subject of a more in-depth review (See Appendix B, Exhibit 5). This review screened the 32 reports to determine whether they met the following criteria:

- Provided number of students tested in American history in TAH classrooms and non-TAH classrooms.

- Provided American history test score data for students in TAH classrooms and non-TAH classrooms.

- Student test score data were reported separately by grade level.

- Student test assessments used to compare student performance from one time point to another were parallel or vertically scaled, e.g. studies that use project developed assessments for pre-test could not use a state assessment for the post-test.

Twelve reports were identified that met the above criteria. The Stage 3 screening further reviewed these 12 studies to determine if they met the following additional criteria:

- Involved student learning that took place in the context of the TAH program.

- Contrasted conditions that varied in terms of use of the TAH program. To qualify, learning outcomes had to be compared to conditions where the TAH program was not implemented.

- Included results from fully implemented TAH projects that delivered their core program components as planned. The types and duration of TAH teacher professional development activities in which teachers were involved varied. For example, some teachers participated in a series of field trips and associated discussions, whereas other TAH activities required that teachers be enrolled in courses for several hours a week over a semester.

- Reported American history learning outcomes that were measured for both TAH and non-TAH groups. An American history learning outcome must have been measured in the same way across the study conditions. A study could not be included if different measures were used for the TAH and non-TAH groups or for the pre-and-post comparisons. The measure had to be a direct measure of learning and not a self-report of learning. Only test score averages could be included in a meta-analysis. Examples of learning outcomes measures that qualified included statewide, standardized tests of American history, scores on project-based tests of American history, performance on NAEP American history items, and performance on student work (in response to American history classroom assignments) scored according to the Newman, Bryk, and Nagaoka (2001) methodology. (Measures of attendance, promotion to the next grade level, grades in American history, percent of students achieving proficiency on statewide tests, and percent correct on statewide measures of American history could not be included in a meta-analysis.)

- Reported sufficient data for effect size calculation or estimation as specified in the guidelines provided by Lipsey and Wilson (2001).

- Used a rigorous design, controlling for relevant pre-program differences in student and teacher characteristics, e.g. pre-program student achievement, pre-program teacher qualifications.

An additional concern about the grantee evaluations is whether project-based assessments, the type of assessment that was most sensitive to TAH effects, were accurately assessing students' achievement of American history. There was insufficient information included in grantees' evaluation reports to enable a review of these project-based assessments or determination of their alignment with state or NAEP measures. More information about the student assessments used by the 12 grantee evaluations included in the final stage of screening, including their reliability and comparability, is included in Appendix B.

The final stage of screening determined that the 12 studies remaining after the first two screening stages met the above criteria with the exception of the use of rigorous designs. Therefore, the studies were deemed to be of insufficient quality to support a meta-analysis.

Exhibit 1 summarizes the characteristics of these evaluations. Only two studies employed a quasi-experimental pretest, posttest design using a treatment and control group; nine studies were posttest only, treatment vs. control; and one study used a one group pretest, posttest design. (Citations for these reports are in Appendix B.) Most of these studies did not use covariates to control or equate the groups of students for prior achievement. A further consideration is that not all of the TAH evaluations took into account differences in teacher backgrounds. Previous research on the TAH program found that participating TAH teachers were more experienced than the average American history teacher (U.S. Department of Education, 2005). If experienced teachers are more likely to participate in TAH programs, this could contribute to the positive effect size. Exhibit 2-2 summarizes the bases for exclusion from a meta-analysis for all 94 studies.

Exhibit 1: Characteristics of 12 Studies in Final Stage of Review

Study Code	Assessment	Study Design	N for each group *	Mean by group	SD for each group	Effect Size Available**	Effect Size Calculated	T/F Statistic	Multiple grades
1	Project Developed	2 Group Pre Post	Yes	Yes	No	Yes	0.15	No	Yes
3	TCAP Statewide Achievement Test in Social Studies	2 Group Pre Post	Yes	Yes	Yes	No**	0.04	Yes	Yes
4	California Standards Test	Post Only T v C	Yes	Yes	Yes	No**	0.18	No	Yes
5	California Standards Test	Post Only T v C	Yes	Yes	Yes	No**	0.32	No	Yes
7	NAEP	Post Only T v C	Yes	Yes	Yes	No**	0.45	Yes	No
8	California Standards Test	Post Only T v C	Yes	Yes	Yes	No**	0.27	No	Yes
9	California Standards Test	Post Only T v C	Yes	Yes	Yes	No**	0.15	Yes	Yes
10	Project Developed	Post Only T v C	Yes	Yes	Yes	No**	0.55	No	Yes
11	Project Developed + Reading and Writing on CT Statewide Test	Post Only T v C	Yes	Yes	No	No**	0.31	Yes	No
12	Student Work/ Newman and Bryk	Post Only T v C	Yes*	Yes	No	Yes	0.00	Yes	Yes
13	TAKS Texas Statewide Social Studies Test	1 Group Pre Post	Yes	Yes	Yes	No**	(-) .188	Yes	Yes
14	PACT	Post Only T v C	Yes	Yes	Yes	No**	(-) .014	No	Yes

Exhibit 2: Bases for Excluding Studies from a Meta-analysis

Primary Reason for Exclusion	Number Excluded	Percentage Excluded
Did not analyze student achievement outcomes in American history	37	40.2%
Did not use a TAH and non-TAH two-group experimental or quasi-experimental design	24	26.1%
Did not use the same measure of student achievement for the pre- and posttest	2	2.2%
Did not report sufficient data for effect size calculation	19	20.7%
Insufficient rigor; insufficient pre-program controls	12	13.0%

Exhibit reads: Of 94 studies reviewed for consideration in a meta-analysis, 37 (40.2 percent) were excluded because they did not analyze student achievement outcomes in American history.

A more limited review of the 2008 Annual Performance Reports of the 2006 grantees indicated that most of these grantees were using single-group designs. In addition, project-developed assessment instruments used for students and teachers were not always thoroughly documented. This suggests that the evaluations of the more recent grantees are also unlikely to be suitable for a meta-analysis.

The weaknesses of the local evaluations of the TAH grantees are a direct result of the many challenges facing local evaluators. In the next section, we summarize those challenges and underscore the point that the lack of rigorous local evaluations was a result of limited resources, real-world constraints, and the needs of projects.

Evaluation Challenges

Local evaluators of the case study sites reported facing a number of challenges in their efforts to conduct outcomes-focused evaluations. Foremost among these was the difficulty of identifying appropriate, valid, and reliable outcome measures. For assessment of students, some evaluators noted that standardized assessments, if administered in their states, were not fully aligned with the focus of the grants. For measurement of teacher content knowledge, nationally or state-validated teacher measures were not available to the grantees in this study. Evaluators developed a variety of project-aligned measures to assess student historical analysis skills, teacher knowledge, and teacher practice. Most of these measures did not undergo formal reliability or validity testing or were not thoroughly documented in evaluation reports. However, a number of project-developed measures are promising and are worthy of further development; these are discussed further below. Overall, the lack of proper assessment tools left local evaluators scrambling to figure out how to measure the contributions of the projects.

Local evaluators were particularly challenged by the difficulty of identifying and recruiting matched comparison groups. Typically, the potential pool of comparison teachers was small and available data were insufficient to determine whether the comparison teachers' backgrounds and previous performance matched those of the treatment teachers. In addition, schoolwide or districtwide dissemination of grant resources potentially resulted in "contamination" of comparison groups. In some regions, the awarding of multiple TAH grants in successive cohorts further limited the number of potential comparison teachers. Even when the local evaluator was able to identify a suitable comparison group, obtaining teacher cooperation was difficult.

Generally, local evaluators lacked strong incentives to motivate comparison teachers to participate in the evaluation.

Local evaluators were also challenged by the needs of the projects and the requests of the project directors. To assist the projects in monitoring their progress, local evaluators administered student and teacher attitude or knowledge surveys, workshop and program satisfaction surveys, and teacher focus groups and interviews. This formative information helped project directors assess implementation fidelity and guide program improvement. Of course, these activities also diverted resources away from more rigorous outcomes evaluations.

Promising Project-based Assessments

Despite these challenges, our case studies revealed some promising evaluation efforts. Several of the case study projects had devoted considerable effort and creativity to designing project-based assessments that were more closely aligned with project goals and focus than standardized tests. This section describes categories of promising alternative approaches that TAH projects have used either instead of or in combination with selected response tests.

Tests of Historical Thinking Skills. Several of the project directors, historians, and evaluators interviewed cited measuring growth in historical thinking as the greatest unmet assessment need. Although selected response tests can be used to measure historical thinking skills, they observed that they do not fully capture the complexity of the subject matter, noting that both the NAEP and the AP American history tests include short and long answer constructed responses in addition to multiple choice test items.

One case study grantee had as its two most important goals the growth of teacher and student historical thinking skills and use of these skills to evaluate primary source documents. The evaluator sought to assess these skills, developing two somewhat similar measures, one for students and one for teachers. The student measure consisted of five questions that were adapted for use as both a pre- and a posttest. Different versions were developed for each of two grade levels. Students were asked to identify and give examples of primary and secondary sources. Students were presented with a primary source (for example, for one grade a Nebraska bill of sale and for another grade an advertisement looking for a runaway slave) and were asked to consider themselves in the role of a historian. They were then asked to write three questions a historian might want to ask to know more about the source. The team developed a simple rubric to grade the responses.

Change in teacher knowledge was assessed using a similar but more sophisticated instrument. Because teachers in the two grades targeted were responsible for teaching different periods in American history, different content knowledge was assessed at each grade level. However, a common subscale was used, as described below:

- Describe what is similar about a series of several matched content pairs (American Colonization Society and Trail of Tears, for example).
- Define primary sources and give three examples.
- Define secondary sources and give two examples.

- Look at primary source documents (such as a map and a cartoon). Write three or four questions that you might ask students about this source.

The rubrics were developed by the historians and evaluator and yielded evidence of teacher growth in historical thinking skills and complexity of responses.

Several other projects have measured historical thinking skills using a document-based question (DBQ) approach. In one example, students at each grade level were given an excerpt from a primary source document and asked to respond in writing to a series of several questions, which range from descriptive to interpretive. One project using this approach validated the assessment by matching the responses to similar constructs measured using multiple-choice questions. In another approach respondents were asked a single question and expected to develop an essay that reflected on the ability to develop a historical argument, provide multiple perspectives, and demonstrate other features of historical thinking. In a third example, teacher reflection journal entries were scored using a holistic rubric.

Assessment of Student Assignments and Lesson Plans. The evaluation of teacher lesson plans, teaching units, or student assignments was another potentially useful form of teacher assessment. Teachers were especially enthusiastic when they received feedback on their lesson plans from historians as well as evaluators. Several sites used a lesson plan evaluation approach grounded in previous work by Newman and Associates (1996) and developed at the National Center on School Restructuring and the Consortium on Chicago School Research. The approach is based on the assumption that when teachers assign certain kinds of rigorous student work, student achievement improves. Assignments are expected to foster students' construction of knowledge and in-depth understanding through elaborated communication. One TAH evaluator developed a lesson plan evaluation rubric that incorporated these constructs as well as indicators of alignment with instructional goals of the project, such as integration of primary sources.

Classroom Observation. Some case study sites used classroom observation both for individualized feedback to teachers and for project evaluation purposes. Observation protocols varied considerably in their goals, structure, content, level of detail, and rigor. Observations might be conducted by the project director, the evaluator, historians, or master teachers. Some observations were highly structured, such as those that included a time log for recording student activities and levels of engagement at five-minute intervals, while others were rated based on a single holistic score. Among the topics evaluated were:

- The use of the historical knowledge and themes covered in the grant.

- The use of teaching strategies covered in the grant.

- Assignment of student work matched to the lesson objectives.

- The use of specific strategies covered in the grant to teach historical thinking skills.

- The use of questioning strategies to identify what is known and what is not, form a hypothesis, and relate specific events to overarching themes.

- The thinking skills required of students during the lesson, often based roughly on Bloom's Taxonomy (Bloom, 1956), which specifies skills such as: information recall, demonstration of understanding, application activities, analysis of information, synthesis, and predictions.

- The levels of student engagement.

- The integration of technology.

In one project the evaluator analyzed the teacher feedback forms using statistical software to determine which historical thinking skills were used most frequently and which levels of cognition were required during the observed activities.

Challenges and Opportunities of Project-based Assessments

The Department has encouraged applicants for TAH grants to include strong evaluation designs in their applications, including measures of student achievement and teacher knowledge. The FY 2010 competition continued the practice of requiring projects to address GPRA measures. GPRA Performance Measure 1 encourages the development of new outcomes measures. As Performance Measure 1 states: "The test or measure will be aligned with the TAH project and at least 50 percent of its questions will come from a validated test of American history."

The promising project-based assessments described here are a response by the 2006 case study grantees to a need for more nuanced measures of teacher and student knowledge gains that may result from TAH projects. However, many program directors and evaluators have noted that the development of alternative assessments requires a level of time, knowledge, and technical expertise that is beyond the ability of individual programs to undertake. Time and expense are required to train scorers and to administer and score assessments. Other challenges include developing grade appropriate prompts, selecting pre- and post-prompts that are at a similar level of difficulty, and developing validity and reliability checks. Without further support, grantee-level evaluators have been unable to take the project-based assessments they have developed to the next level of refinement and validation. However, many of the project-based assessments discussed provide frameworks, which could potentially be adapted to varied contexts and content and are worthy of further exploration and support.

Conclusions

Review of grantee evaluations for this study found that TAH evaluations should be more rigorous if they are to be used to draw conclusions about the overall impact of the TAH program on student achievement. The screening of 94 final evaluation reports of 2004 grantees for possible inclusion in the meta-analysis revealed that the great majority of evaluation reports lacked detailed information about the sample, design, and statistical effects. Moreover, most local evaluations lacked adequate controls for differences in teacher qualifications and most did not control for previous student achievement.

The Department has made a concerted effort to determine the contributions of the TAH program to student achievement in American history. Some of these efforts have focused on encouraging local evaluators to carry out rigorous research designs. This approach has not yet been successful. Local evaluators are struggling to find or develop appropriate assessment tools and to fully implement rigorous experimental designs. The implications of these challenges are discussed in the final chapter of the report.

Chapter 4
Strengths and Challenges of TAH Implementation

A major goal of the study is improved understanding of those elements of Teaching American History projects that have the greatest potential to produce positive achievement outcomes. This chapter presents results of case study research, addressing the key questions:

- What are strengths of TAH grantees' program designs and implementation?

- What are major challenges that impede program implementation?

Researchers conducted case studies of 16 TAH grantees of the 2006 funding cohort in order to identify and describe project practices most likely to lead to gains in teacher knowledge or student achievement.[8] Case studies entailed in-depth site visits with teacher and staff interviews, and—at most sites—observations of professional development.

The selection of case study sites focused on identification of grantees who reported greater than average improvements in teacher content knowledge or student test scores, for comparison to more typically performing grantees. Four grantees with improvements in students' state American history test scores were compared to four grantees who did not exhibit gains; four grantees with improvements in teachers' content knowledge (based on data in Annual Performance Reports) were compared to four grantees who did not provide evidence of such gains. The outcomes data used to select and categorize grantees had several limitations. Factors other than the TAH program might have been responsible for changes in students' performance in American history over the course of the grant. Outcomes data on teacher knowledge were based on grantees' self-report; researchers could not confirm the reliability or comparability of the measures. Finally, outcomes data used for case study selection were 2008 data and therefore represented two years—rather than the full three years—of grantee performance.

Given these limitations, researchers used the research literature on effective practices in K–12 teacher professional development to set benchmarks for identification of promising practices among the case study sites.

Key findings of the case study research include:

- No systematic differences were found in practices of grantees with stronger and weaker outcomes.

- TAH projects aligned their practices with research-based professional development approaches through the following practices: professional development that balanced content and pedagogy; the employment of project directors who coordinated and managed this balance; the selection of partners with both content expertise and responsiveness to teachers' needs; clear expectations and feedback for teachers; and the creation of teacher learning communities and other forms of teacher outreach. Most projects were not implemented schoolwide, and support from district and school administrators was uneven.

---

[8] A more detailed discussion of case study design, selection methods, and limitations of the selection process is provided in Appendix A.

- A persistent challenge facing TAH grantees was the recruitment of teachers most in need of improvement. Grantees used a wide variety of strategies to recruit teachers. Among these strategies were conducting in-person outreach meetings at schools to recruit teachers directly and offering different levels of commitment and options for participation so that teachers could tailor participation to their schedules and needs.

The case study sites cannot be considered representative of all grantees, and findings cannot be generalized beyond these 16 sites.

Participants' Views of the Projects

During the site visits researchers interviewed close to 150 individuals, including teachers, project directors, partners, evaluators, master teachers, and other staff. Almost universally, respondents reported that participation in the TAH programs significantly increased teachers' content knowledge in American history. Teachers frequently lauded the professional development as "the best in-service I've ever had." Many teachers echoed the response of this teacher who observed, "What I learned in the three or four years I've been here, from the professors that come and talk to us, outweighed what I learned in college, by far." Teachers and project partners noted that, in general, history and social science teachers have far less access to professional development opportunities than do teachers of reading, mathematics, or science. They noted that the TAH program helped redress this imbalance, and the quality of the presentations by historians "reenergized" many history teachers who were eager for new knowledge and skills.

When asked how they measured the grant's success, most teachers and project staff focused on improvements in teachers' content knowledge and teaching skills and on students' classroom engagement and understanding of history. As one teacher said, "I'd like to think that I became more excited and passionate about history, and that translates to students. I don't know how to quantify that." Although a few teachers were aware of improvements in their students' scores on standardized American history tests and attributed those changes to the project, many teachers did not focus closely on test scores as a measure of the grant's success.

Teachers emphasized their increased access to primary sources (through presentations by historians, field trips to historical sites, the discovery of new websites and their own research). They reported a growing sophistication in how to integrate primary sources into instruction and remarked on the resulting benefits as a means to convey the many ways history can be interpreted and for making history more exciting and real to their students. Some noted that the "multisensory" nature of the primary sources provided through their projects—including written texts such as letters, speeches, and diaries as well as photographs, paintings, maps, political cartoons, interview tapes, and music—provided a richer historical context, facilitated critical thinking, helped students to compare and contrast themes, and evoked personal connections with history. This improved students' memory of historical facts and assisted struggling readers in framing and understanding difficult texts. Teachers also noted that due to their application of new techniques for encouraging historical thinking, students were now more likely to ask questions, to see history as a process of inquiry, and to take the initiative to pursue answers to their own history questions on the Internet. As one teacher explained:

My teaching, because of this grant, has dramatically improved. I went from someone who was more of teaching to the test, to really focusing on critical thinking…. I have changed my philosophy of teaching for the better....

Strengths and Challenges of TAH Professional Development

Despite efforts to identify practices associated with positive outcomes, no systematic differences were found in practices of grantees with stronger and weaker outcomes. As noted above, outcomes data used to compare and select sites had a number of limitations. In addition, the multifaceted nature of the programs, the complexity of the data, and the variation within the categories may have confounded any relationships between practices and outcomes given the small sample of case study sites.

Nevertheless, the case study research documented ways in which TAH projects were able to align their practices with principles of high-quality professional development as defined in the research literature and by experts within the field. In this chapter, we elaborate on how the case study sites implemented, adapted, or struggled with each of the following elements of high-quality professional development:

- Balancing of strategies to build teachers' content knowledge and strengthen their pedagogical skills.

- Employing project directors with skills in project management and in the blending of history content and pedagogy.

- Building partnerships with organizations rich in historical resources and expertise, and flexible enough to adapt to the needs of the teachers they serve.

- Obtaining commitment and support from district and school leaders and aligning TAH programs with district priorities.

- Communicating clear goals and expectations for teachers and their project work and providing ongoing feedback.

- Creating teacher learning communities, including outreach and dissemination to teachers who did not participate in TAH events.

- Recruiting sufficient numbers of American history teachers to meet participation targets for TAH activities, including teachers most in need of improvement.

Balanced efforts to build teachers' content knowledge and strengthen their pedagogical skills

The TAH grant program has long emphasized the need to develop the content knowledge of history teachers in this country. Previous research on history teacher preparation has shown that teachers often do not know how to practice the discipline themselves and therefore lack the capacity to pass critical knowledge and skills on to their students (Bohan and Davis 1998; Seixas 1998). Teachers need both depth and breadth in their knowledge of American history in order to teach to challenging academic standards. However, teachers must also know how to integrate this new knowledge with high-quality teaching practices if they are to impart the knowledge to

students. As one TAH evaluator pointed out, optimal programs offer a "seamless mix of the content and how to teach it." A balanced approach to teacher preparation allows for multiple cycles of presentation, assimilation, and reflection on new knowledge and how to teach it (Kubitskey and Fishman 2006). Features of professional development that can help achieve this balance include collective work, such as study groups, lesson planning, and other opportunities to prepare for classroom activities with colleagues (Penuel, Frank and Krause 2006; Darling-Hammond and McLaughlin 1995; Kubitskey and Fishman 2006), "reform" activities such as working with mentors or master teachers, and active learning in which teachers learn to do the work of historians through collaborative research (Gess-Newsome 1999; Wineburg 2001; Stearns, Seixas and Wineburg 2000).

Although the TAH program has always placed an emphasis on building teachers' content knowledge, many of the case study grantees chose to balance this goal with improving the instructional practice of participating teachers by providing them with new teaching strategies, lesson plans, and classroom materials. This balanced approach was viewed by the project directors and teachers as a way to sustain the interest and motivation of teachers, provide teachers with tools for differentiating instruction for students at different grade levels and with varied backgrounds, increase student engagement, and ultimately improve student achievement in American history. Additional goals of some grantees were to help teachers align their teaching with state standards in American history and to improve students' scores on state American history exams and Advanced Placement exams. The case study grantees' approach to achieving this balance varied. Several grantees split their summer workshops into two sessions, with the first half of the day devoted to a historian's lecture and the second half of the day focused on showing teachers how to apply this information in their classrooms. Some grantees brought in professional development providers to conduct workshops specifically devoted to pedagogy; others used master teachers to model lessons and work directly with teachers on lesson plans and activities that incorporated historical knowledge, resources, and skills that they were gaining through the grant. The following are illustrations of grantee strategies for providing this balance.

Varied Modes and Timing of Professional Development. All but two of the 16 case study sites offered summer professional development institutes of one or two weeks in length. The institutes either focused on a specific historical period, such as the colonial era, or a theme, such as conflict and consensus building. All sites offered school-year professional development as well. These school-year activities varied widely across the projects. Lectures by historians were provided throughout the year. Several programs offered extended research action programs researching local historical sites. In two programs, teachers learned about collecting oral histories, conducted local interviews, and planned lessons around the oral histories. Many programs offered after-school book study groups, often facilitated by historians. In some cases, teachers' journal reflections on their reading were shared on a projectwide discussion board. Most programs offered occasional field trips on weekends. These trips might be developed around a historical theme or include a trip to a local archive to conduct research. A series of Saturday sessions might be offered on lesson planning or on specific topics such as how to develop document-based questions (DBQs) for student assessment, or how to develop PowerPoint presentations based on primary source documents. Some teachers also attended local and national conferences such as the Organization of American Historians, the American Historical Association, and the National Council for the Social Studies, and reported back to their colleagues. Most sites required that participants make a commitment to attend a minimum number of events or complete a minimum

number of hours of TAH professional development; this requirement usually ensured that participants experienced a mix of lectures and more "hands-on" activities.

Teacher Resource File. In one of the single-district case study programs, content-rich lectures and seminars by scholars were consistently accompanied by sessions on incorporating instructional strategies and resources related to the content covered. In addition, participating teachers were provided with a cohesive and carefully planned "resource file" designed to support classroom integration of the content and pedagogy learned through professional development events. It was evident that teachers valued and used the many materials they had acquired through the grant, as well as the resources to which they were directed in pedagogy sessions. Among the resource file materials noted by teachers were:

- Teacher binder for activity notes and materials.

- Scholarly books for teacher reference.

- Student-friendly books (especially those with primary source material).

- Technology and visual pieces, such as video clips, oversized historical photos, and primary source kits.

- Local materials and primary sources when available.

- Teacher's choice of a classroom set of primary source materials on a specific topic.

During interviews, teachers gave examples of referencing these materials, regularly incorporating them in lessons, and sharing them with other teachers, emphasizing that they "don't gather dust on the shelves."Another teacher described her classroom library:

> *"I cleaned everything out this fall, reorganized and realized that so much of what I had has come from this program. I can honestly say that I've been able to use 90 percent of it."*

Several teachers especially valued and frequently mentioned the oversized historical photos as a tool for engaging students and teaching primary source analysis. Project staff mentioned the value of providing scholarly books for teachers' own reference, and teachers mentioned using them for research and ideas for lessons. In two other projects associated with a national partner, "History in a Box" packages were made available for loan. These nationally developed materials contained a collection of multimedia resources developed around a historical period, such as Westward Expansion, or a famous person, such as Abraham Lincoln.

Mentor Teachers. In another grant, mentor teachers helped ensure the balance of content and pedagogy. Five mentor teachers were selected based on their prior leadership and mentorship experience and their qualifications in history education. The grant relied on these mentor teachers to provide advice on aligning the content-focused professional development delivered by historians with the state standards and—more importantly—to work with the teachers to incorporate what they were learning through the grant into lesson plans that would meet the standards. The mentor teachers were involved in the project planning team as well. Historian partners receive feedback from the mentor teachers on how to make history content engaging and useful to teachers. The mentor teachers were critical partners, as they provided the "pedagogy": they worked with teachers in grade-level groups at the end of professional development sessions to help them apply what they had learned, link the history content to district standards, and

develop lesson plans. Teacher feedback suggested that the mentor teachers contributed enormously to the pedagogical applications of the historical content, but teachers also reported that more ongoing contact with the mentors, especially in-between formal professional development sessions, would have been even more helpful. Based on this experience, a greater emphasis on ongoing mentoring has been incorporated into more recent grants. Overall 9 of the 16 case study sites employed mentors or master teachers.

An Evolving Emphasis on Pedagogy. For one grantee and its university partner, the balance between improving teachers' knowledge of American history and improving their teaching skills evolved over the life of the grant. Having had their first application rejected for not focusing enough on history content, their successful grant application emphasized history content knowledge almost exclusively. The first summer institute was comprised of a series of all-day lectures by historians. Teachers heard six and one-half hours of lecture on the colonization of North America. Critical feedback from participants led to major changes in the second year of the grant. The second summer institute included a mix of lectures, walking tours, discussion groups, and lesson planning. The institute began with a classroom simulation—a debate over concepts of freedom in the Atlantic world before 1800. In addition, historians were paired with master teachers in an effort to ensure that the summer institute and the periodic workshops included both the presentation of rich historical content and practical ideas on how to use that knowledge in the classroom.

Thinking Like a Historian: Analysis of Primary Sources. Teachers valued historians' lectures not only for their content but also for instilling in them a better understanding of history as a specialized form of inquiry based on the analysis of historical evidence. Using primary source artifacts as well as the work of other historians, a lecturer might model the process of forming a hypothesis about a historical event or topic, comparing and contrasting different interpretations and reaching a new or original conclusion. Through this process, teachers increased their understanding of the many ways history can be interpreted. Some teachers observed that they had not previously realized how much their own course textbook left out and began to see the value of relying on other sources in addition to the textbook.

Teachers reported that they found they could transmit this understanding to their students, especially if given concrete strategies and materials to use in the classroom. Professional development that focused on interpretation of primary sources offered a number of opportunities for combining content and pedagogy. Participants learned specific teaching strategies, such as how to:

- Use primary sources to set the historical background or context.

- Select short student-friendly, age-appropriate sources, such as excerpts from a document, photographs, or songs.

- Group primary sources by themes.

- Use photographs they had taken themselves during field trips.

- Develop a set of questions to promote specific higher-order historical thinking skills such as how to see a historical event through the eyes of different groups, understand patterns, establish causes and effects, or understand the significance of an event within a broader context.

- Connect primary sources to present-day issues relevant to students.

- Teach students how to collect their own primary sources.

The example below illustrates how both auditory (music) and visual (musical event program covers) primary sources could be used to impart history content, historical analysis skills, and pedagogical skills.

An Example of Using Primary Sources to Convey Both Content and Pedagogy

In one TAH project, historians from the Smithsonian Institution and local universities taught teachers to analyze musical pieces and program covers for musical events from the mid-19th century to better understand—and teach—culture and race relations during the period. The lecturers modeled a "think aloud" process, verbalizing what they were thinking as they listened to the music and viewed the illustrations, thus demonstrating how a historian might evaluate the "source" artifact and use "close reading" to analyze, question and interpret the artifact. Using this approach, the historians communicated their own extensive knowledge of the topic while at the same time modeling how teachers might identify the text and subtext of visual and musical artifacts with students.

Integrating the Use of Technology. Most projects used technology as a tool to blend content and pedagogy. Teachers commonly reported that they increased their use of technology in the classroom as a result of the grant. In one project the two most common technological tools mentioned were podcasts and wikis. One high school history teacher, for example, developed a wiki for a unit on slavery in the American colonies. The wiki was an online information source that let the students, "click on the links." In a later unit, students were asked to create their own "wikispace" about a topic related to Westward Expansion. Another teacher used a wiki to adapt his instruction for English learners. For each chapter in the textbook he downloaded an audio recording. "They can listen to it as they read, and for second language learners that is huge," he noted.

Most of the case study sites had project websites and uploaded teacher-developed lesson plans. Sites also provided links to national organizations that have developed materials for teachers. Project directors noted that, since the inception of the TAH grant program, there has been a significant increase in the online resources for teachers provided by national history organizations. Several national organizations serve as partners in TAH programs and initially developed the materials as part of their project work, later expanding to a national audience. In one project, for example, site visitors observed a workshop given by members of one of the largest national history organizations. TAH teachers were asked to provide feedback on the relevancy and usefulness of the field test version of a lesson planning tool that provides multimedia lesson plan materials on a wide variety of historical periods. Teachers can quickly browse the materials according to various subtopics, select their grade level, specify whether they need a 30-, 45-, or 60-minute lesson and with the "click of a button" produce a lesson plan. Teachers at this same site also had access to lesson plans and curriculum correlations through an online media streaming service made available by a local television station. As one teacher noted:

> *"I would have learned the content and skills without TAH, but it would have taken longer. TAH was a shortcut. I improved my content delivery, improved my lesson, and made better use of technology. This was a chance to get it all quick. I benefitted a lot because I learned so many things. I use technology almost every day now."*

Field Trips. Visits to local museums, historical sites, and archives were a feature of every program visited for the case study; most teachers reported that these first-hand experiences significantly deepened their history content knowledge and pedagogy. Several teachers noted that the field trips inspired their interest in local history. "I want to be able to not only talk out of a book but to have a more hands-on understanding," one teacher observed. Arrangements were often made for highly knowledgeable tour guides, archivists, or historians to work with the teachers, and special "behind the scenes" tours were set up. Teachers reported that being treated as a historian elevated their concept of themselves as professionals. As one teacher noted, "One of the things that I love is that teachers feel really respected." A project director noted that during the field trips the teachers often were "validated in ways they don't get in other aspects of their careers." Teachers not only learned from the tour guides and historians that accompanied them, but also from each other, especially about how to use the information in teaching. One teacher observed, "There were [other teachers] who just seemed to have a lot of information and high level of expertise…. [I learned by] talking to colleagues about how they have used the information…."

Strong project directors with skills in project management and in the blending of history content and pedagogy

As research on educational leadership has shown (Bennis and Nanus, 1985; Duttweiler and Hord 1987), the person on the "front line" of educational change needs to be both a logistics manager and an instructional leader with the skills to execute the format and progression of activities. Leaders must bring in and motivate outstanding experts and evaluators, work with a team to set focused, transparent goals, and implement ongoing program improvements using feedback from all stakeholders. In the TAH case study sites, project directors were able to leverage their skills and knowledge and the expertise of local staff, partners, and evaluators to plan and implement a team-based approach. In at least one instance, the project director was the key to keeping the project moving forward in the face of obstacles presented by the district's finance office that delayed approval to conduct grant activities. Following through with the participants to gather feedback was also highly important, as one participant reported:

> *"[The project director] is so good at getting back to you and planning. He spends an awful lot of time getting the best of the best for his people. I think that the very small group [running the grant] is essential because the money goes not to administering the grant but to the people participating in it, and I think that that's a big deal."*

Project Directors With History Teaching Experience. The value of teachers as professional development leaders is supported by research findings (Lieberman and McLaughlin 1992; Schulman 1987) that current or former classroom teachers are often perceived to be more credible and to provide professional development that is more meaningful to teachers. Experience within the district culture can provide insights that allow a project director to create coherence within the project and alignment of teachers' goals, project goals and district goals (Coburn 2004) and to better overcome district policy and management hurdles.

Many of the strongest project directors were well-respected current or former history or social studies teachers with many years of experience in the district who had been promoted to become department chairs or district-level curriculum specialists. Some teachers, in fact, reported that

they were attracted to the program based on the reputation of the project director. As one teacher said, "I knew anything [the project director] was involved with would be great." In addition, as history and social studies teachers themselves, they are perceived to be more "credible"—"The fact that [the project director] is a teacher as well helps. She understands what teachers want and what teachers need." Finally, strong project leaders were able to communicate a level of commitment and love of the field, as in the case of this project director:

> *"She is really on top of things. Part of it is that she loves history. Teachers share her enthusiasm and it is generated by her knowledge of all these museums. She has a strong knowledge of what is out there."*

Project Director Guidance of History Partners. Participants valued project leaders with in-depth knowledge of how to select and guide the expert historians. Strong project leaders screened history experts in advance to make sure their presentations included information about how to translate history knowledge into classroom activities, or how this knowledge related to district content standards. Throughout the course of the grant, these leaders were able to maintain a strong working relationship with all partners, which helped to facilitate communication and decision-making from initial planning through the final stages of implementation.

Project Director Response to Constructive Feedback. Teachers also appreciated project directors who were able to take constructive feedback from project participants and include it in subsequent project offerings. Speaking about her project director's ability to incorporate the opinions of project participants, one veteran TAH teacher noted that, when communicating with her director, "Your feedback is always listened to; if you ask for something it's there the next year. Many teachers have had a great experience with the grant basically because of [the director's] involvement." Project directors such as this one had a "very good idea of the big picture of this grant, all the way down to the smallest details of the grant." Many worked closely with the project evaluator to review data collected after project activities and through focus groups and results of teacher content knowledge assessments. The on-going changes made by project directors included a stronger blending of content and pedagogy, the development of activities tailored to teachers at different grade levels, and the offering of varied levels of teacher involvement based on their other professional commitments.

Project directors were reviewed less favorably by participants when they were not accessible to participants, when they were less directly involved with the professional development delivery (viewing their role more narrowly as managers), and when they failed to clearly communicate about the project. In some cases, the grant was plagued by turnover of project directors. As one project manager, who experienced turnover of almost all the original team members noted, "The original vision has been somewhat lost over the years," including the intent to tap into the community and the historical character of the local region.

Partnerships with organizations rich in historical resources and expertise, and flexible enough to adapt to the needs of the teachers they serve

The TAH program requires that grantees have commitments from partner organizations capable of delivering in-depth history content to teachers. Although substantial work has been done examining the role of the district in professional development (Andrews 2003; Snipes et al. 2003; Elmore 2004), more limited research has addressed the role of community partners and postsecondary institutions in providing effective in-service professional development (Desimone,

Garet, Birman, Porter and Yoon 2003; Watson and Fullan 1991; Teitel 1994; Tichenor, Lovell, Haugaard and Hutchinson 2008). As Desimone and her colleagues point out, much of this research has been related to the professional development of mathematics and science teachers (Desimone, Garet, Birman, Porter and Yoon 2003). For example, as part of the large Eisenhower Professional Development Program, researchers examined the management and implementation strategies provided by postsecondary institutions to determine what contributed to high quality in-service teacher professional development in mathematics and science. They found empirical support for the concept that aligning professional development to standards and assessments, implementing continuous improvement efforts, and ensuring coordination between postsecondary institutions and districts improved the quality of professional development. Some studies have examined the challenges faced by partnerships, such as integrating cultures, territory disputes and dealing with funding issues (Teitel 1994). Others have focused on how university partnerships can help make school improvement processes more coordinated and focused (Watson and Fullan 1991; Bell, Brandon and Weinhold 2007) and break through the physical and intellectual isolation of teachers (Carver 2008). A more limited number of studies have examined the role of museums (Hooper-Greenhill 2004) in teacher professional development.

Several authors have recently described the role of university and community partners within TAH projects (Woestman 2009; Knupfer 2009) and reflected on the conditions for productive collaborations and the benefits of TAH participation for professors, such as increasing knowledge of pedagogy and the needs of teachers (Apt-Perkins 2009).

Case study participants reported that strong partnerships were integral to the successful implementation of the TAH programs. Access to highly qualified historians was cited as the most important benefit provided by the partners. Project staff and participants noted that effective lecturers were not only well-versed in their content area but also were able to model analytical processes for thinking about a historical topic from multiple perspectives.

In most case study sites, an institution of higher education or a national history organization was the lead partner. Typically, a faculty historian with the lead organization served as an academic advisor. Optimally, the academic advisor provided continuity within the project by participating in most activities and coordinating the ongoing professional development offered by other historians brought in for their expertise in specific topics. Other valued contributions of partners were: advising master teachers, selecting reading material for teachers, observing classroom teaching, reviewing lesson plans, and assisting in the development of teacher and student content knowledge tests. In about a quarter of the projects, a university, national nonprofit history organization, or community development agency also played a leading role in project management. Other partners included state historical societies, state humanities councils, local public broadcasting organizations, local television channels, the National Park Service, art museums, nonprofit legal rights organizations, nonprofit teacher training organizations, for-profit curriculum development institutes offering commercial curriculum, and individual consultants.

Partners contributed to projects in widely varying ways. Historians who delivered professional development were praised by participants at almost all sites. Project staff particularly valued partners who were flexible and responsive to teachers' needs. The richness of the mix of partners and the coordination of their various contributions varied. This led to differences in how well the projects integrated historical content with useful guidance on teaching practice.

Partners From Departments of History, Education, and Civic Education. At one well-developed and comprehensive partnership, three branches of the same state university were integrally involved in planning and implementation. Representatives from the university history department provided rich content expertise; they were well supported by faculty at the college of education who had strong skills in applying the knowledge in the classroom. An institution on campus that provides programming and scholarships related to civic education contributed to the grant by providing its facilities on the university campus, as well as by providing material resources and access to their network of scholars. They actively shared what they learned with other TAH grantees in the state, thus establishing a network to enhance the professional development of all the grantees in the state.

Partner Support in Research on Local History. In the case of one four-district urban project, the lead partner, a community development agency, established partnerships with a number of local historical sites. At each site they arranged for historians to be available to provide specialized behind-the-scenes tours linked to the historical topics that were the focus of the professional development. A strong partnership with the urban library system then facilitated the teachers' engagement in original research on a topic of their choice; librarians worked with teachers to produce significant local historical research. For example, one teacher wondered about the fate of Native American children after a major 17th-century massacre in the local area. Through her research in the archives of the public library, she discovered newspaper advertisements offering Native American children for sale, a practice not widely associated with New England. Her research project, supported by multiple historians and archivists, led her to a new approach to using primary sources in teaching history to her students.

Some programs lacked access to such strong partnerships. One project suffered due to the lack of a university with a history education department within its largely rural region.

Some historians who worked with the case study sites noted that the overall level of collaboration established between colleges and universities and public education facilitated by TAH funds is unprecedented in social studies education.

Commitment from district leaders and alignment with district priorities

District support for professional learning and development has long been identified as a key component of improving student performance, as noted by Andrew (2003). Evidence suggests that school districts need to use a large and coordinated repertoire of strategies for staff at all levels in order to improve student achievement (Snipes et al. 2002). Numerous studies have focused on the perceived and actual leadership characteristics and actions of school superintendents in promoting professional development (Peterson 1999) and the role of professional development in districtwide reform (Elmore 2004; Resnick and Glennan 2002).

The initial impetus behind TAH projects at the case study sites often came from a district leader such as a superintendent or assistant superintendent who recognized a need in the district for more teacher training in American history. But interviews with project staff and teachers suggested that ongoing district and school administrator involvement with the TAH program was often limited to passive, hands-off support for teachers to participate in the professional development. As reported by a number of teachers and project leaders, history and social science are a low priority in many districts given the emphasis on reading and mathematics in accountability testing. As a result, obtaining the strong commitment of all district and school

leaders was challenging for some project directors, particularly for grantees engaged in improving American history instruction in multiple districts.

Problems Due to Lack of District Support. At some sites, teachers did not find themselves to be impeded in their grant participation by this lack of official involvement at district and school levels. These teachers, who were often from small or isolated districts or schools, enjoyed the opportunity the grants provided to connect with history teachers outside of their districts and to pursue study of personal interest not specifically related to district requirements. However, in other sites, the lack of district and school support meant that district officials and principals were reluctant to allow teachers to be released from their classrooms or other school and district obligations, such as district-mandated professional development, to attend TAH opportunities. Further, because district and school support were needed to encourage ongoing teacher collaboration and diffusion to nonparticipants, benefits of the grant are more likely to fade in the absence of this support.

Involving Superintendents. In a small number of grants, project directors were successful in building relationships with superintendents and aligning grants with other district priorities. District support lent legitimacy to the projects and helped them run more smoothly. The principals in one of the districts initially balked at releasing teachers from school-based professional development days to conduct research for the grant, which created a conflict between the principals and the project director. As a solution, the superintendent offered to pay for substitutes for all the participating teachers to allow teachers to attend both the TAH program and the school-based professional development. Another grant benefited from a cross-district advisory committee. Superintendents from participating districts met regularly to discuss grant programming and implementation issues. By continuing to monitor the grant's progress, these leaders were able to connect TAH programming with other district priorities, such as writing.

Alignment with State Standards. Another pair of grants exhibited moderately strong district relationships and a focus on alignment with state standards. In these grants, professional development activities were designed in part to assist teachers in developing lesson plans well-aligned with state standards. District leaders were also more likely than elsewhere to be actively involved in the planning and development of the projects. It may be that circumstances within these two states, such as fully developed statewide history standards, an emphasis on teaching to standards, and regional entities based on strong district partnerships, created a favorable context for developing district support for the grants.

Noteworthy grantee strategies were those that combined strong partnerships, balanced content and pedagogy, and linkages to state or district standards. The example on the following page illustrates how partners of one project created an opportunity for teachers' research on local history that was in turn used by teachers to create a new curriculum unit that ultimately led to gains in students' attainment of standards.

**An Example of Strong Partnerships Leading to Standards-based Curriculum
Using Local Sources**

In a site located in a major urban area, a number of historians from various universities, as well as a librarian and a local representative from the National Park Service, joined the partnership. The historians urged the team to adopt a project-based model using local primary sources. By doing original research, they argued, teachers would better understand the work and thinking processes of historians. The partners trained the teachers in how to conduct archival research and locate primary source documents about their local area. The teachers began to develop a multidisciplinary project-based unit about a local historical landmark—a large 18th-century factory on the edge of the town center. As they conducted their research they learned that the factory produced "cutting edge" technology for its time. They discovered it was founded by a colorful entrepreneur whose story had been all but forgotten. Working side-by-side with the historians, the teachers devoted many hours during the summer institute to documenting the history of the factory and developing lessons for their students to begin in the fall.

Once the school year began, other teachers became involved, and teachers worked together to develop a curriculum unit. Gradually they created a unit that combined social studies and science in a lesson sequence targeting state standards on which the school's students had been performing poorly. The unit included both an analysis of the historical context surrounding the site and an exploration of the factory's mechanical operation in its heyday. It culminated with a field trip to the factory. The unit was very successful, with the teachers enthusiastically describing the student growth that they observed. Not only did students and the school gain attention from the local press, but students outperformed other students in the district on standardized tests.

Establishing clear goals and expectations for teachers, with ongoing expert feedback

Hallmarks of successful professional development initiatives are clear goal-setting and monitoring of progress toward goals (Gutsky 2003; Desimone et al. 2002; Haskell 1999), a carefully constructed theory of teacher learning and change (Richardson and Placier 2001; Ball and Cohen 1999), and models and materials based on a well-defined and valid theory of action (Hiebert and Grouws 2007; Rossi, Lipsey and Freeman 2004). TAH teachers and project directors at the study sites reported that project success was related to the establishment of similar practices, including a common vision of teacher change and a clear theory of action that aligned project activities with expectations for teachers and guided teachers on meeting these expectations. Respondents reported good results from a process that included: (a) setting clear expectations that teachers produce lesson plans, curriculum units, or independent research products; (b) ensuring follow-through on completion of these products; and (c) providing feedback on these products from historians, lead teachers, or other experts.

Structured Teacher Requirements and Feedback. In one site, participating teachers were asked to sign Memoranda of Understanding that clearly outlined the project goals and expectations that teachers were required to fulfill in order to receive in-service credits, graduate credits, and a teacher stipend. Each day of the summer institute began with a lecture and discussion sessions led by the academic director (a local university historian) or one of the pre-eminent historians he invited. In the afternoon, the group was broken up by grade level. Lead teachers modeled lessons based on the morning's content, and teachers began conducting independent research with the support of the academic director. During the school year, activities included a mix of lectures, lesson planning workshops, book study groups facilitated by the historians, weekend field trips, and Saturday workshops on archival research. Teachers were required to keep reflection journals, excerpts of which were shared on a project discussion board. The academic director, lead teachers, and the evaluator visited the classrooms three times each

year. They used a structured protocol and rubrics for observation and met with teachers to provide feedback. Teachers also received ongoing feedback on interim and final drafts of their original research projects and accompanying lesson plans. Their final presentations were videotaped and the lesson plans (linked to the new district standards) were posted on the project website.

Requirements for Lesson Plan Development. Twelve of the 16 case study sites required teachers to develop lesson plans or units of study as part of their TAH participation. In some cases teachers were expected to conduct original research. Drafts were reviewed by the program director, master teachers, or historians. Teachers were observed teaching the lesson. Presentations based on the lesson plan were then made to colleagues who offered suggestions or considered ways to adapt the lesson for other grade levels or contexts. In some projects the final products were evaluated formally as part of the overall program evaluation process. In other cases the production of lesson plans was a more informal requirement.

Keeping Projects on Track. At the project level, frequent and ongoing meetings to make mid-course corrections to meet the goals were also important. Many successful program teams carefully reviewed responses to teacher surveys collected after major activities and used these to plan changes. For example, one successful program hired grade level specialists for middle and high school teachers when it was found that existing activities did not meet the needs of teachers from different grade levels.

Among the projects that were less successful, the goals of the projects were less transparent and expectations of teachers were limited. A lack of follow-through for the completion of products such as teacher lesson plans and a lack of feedback on the success of the work products resulted in inferior or partially completed work. These problems were exacerbated when there was a high degree of turnover among key staff, especially the project leader, in which case the original "vision" and goals for the project were lost or diluted. In some cases, field trips appeared to be only loosely connected to project goals; teachers commented that there were missed opportunities to reflect upon and consolidate what they had learned from the travel or to develop products such as lesson plans based on the field trips.

Continuity with partners also made a difference in the extent of feedback teachers received. For example, when partners were located at a distance from the project and made infrequent visits for guest lectures, there were fewer opportunities for follow-through and feedback.

Teacher learning communities, including outreach and dissemination to teachers who did not participate in TAH events

A mounting body of evidence supports the benefits of teacher engagement in professional learning communities or networks of information exchange and collaboration. Learning communities provide teachers with opportunities for shared learning, reflection, and problem-solving and allow them to construct knowledge based on what they know about their students' learning and evidence of their progress (McLaughlin and Talbert 2006). There is also evidence that networks of teachers can help sustain teacher motivation (Lieberman and McLaughlin 1992). In the large-scale study of the Eisenhower Professional Development Program (Garet et al. 2001) researchers also found that activities that encouraged professional communication among teachers had a substantial positive effect on enhanced knowledge and skills, as well as on changes in teaching practices. A five-year study by Newman and Wehlage (1995), based on 24

restructured public schools, found that a professional community was one salient characteristic of those schools most successful with improving student achievement. Finally, using data from the National Education Longitudinal Study of 1988, researchers conducted three studies that have consistently shown that teacher communities have a positive effect on student achievement gains (Lee and Smith 1995, 1996; Lee, Smith and Croninger 1997 as cited in McLaughlin and Talbert 2006).

Across the TAH case study sites, a variety of informal and formal collaborations or "teacher learning communities" were in place for participating teachers. Some projects also developed more widespread networks for dissemination and sharing with nonparticipants. The structure and communication modes for teacher networks varied greatly. Some grants required participating teachers to plan and conduct staff development events for nonparticipants in their schools or districts; others shared lesson plans via websites and CDs; others focused primarily on sharing and collaboration among the core project participants.

Teacher networking and collaboration contributed to the grants' penetration, participant commitment, and sustainability. In regional grants serving smaller, more isolated schools and districts, history teachers with few colleagues on-site (in some cases the only American history teachers in their schools) became members of a new community of colleagues who reinforced learning, provided opportunities for collaboration, and shared resources and lesson plans online or in occasional in-person meetings. When networks were developed within schools or districts, they strengthened the schoolwide or districtwide commitment to the new teaching practices or curricula and potentially magnified the impact of the grant on student achievement in the school or district. Networks and learning communities, even if limited to the core participants, were expected to outlive the life of the grant and therefore help sustain the new teaching ideas and practices resulting from the grant. Encouraging or requiring participating teachers to share knowledge and skills gained through the project with nonparticipating teachers was a promising, cost-effective strategy used by some grantees to extend the grant's penetration throughout the districts and to reach teachers who were unable or unwilling to participate in the core activities.

Technology and Rural Teacher Networks. Within one grant that included multiple small rural districts, technology was both an in-class teaching tool and a networking tool among teachers. In one interview, a teacher indicated he used Twitter, a social networking site, to request ideas for a lesson. Within minutes, participating TAH teachers from across the region responded with several ideas of lessons they had delivered, suggestions for activities, and online resources. Because the grant serviced teachers from rural areas that in some case contained as few as one or two history teachers, the development of a regional network via technology became a highly valued component of the grant.

Strong Districtwide Participation. In one single-district grant, teachers and stakeholders spoke at length about the overwhelming success of the network (both social and professional) that resulted from the grant. The positive group dynamic clearly contributed to teachers' ongoing participation and engagement. Possible characteristics promoting relationship-building were strong leadership, regular pedagogy sessions with time for teachers to work together, and opening up selected activities to all American history teachers throughout the district, rather than limiting all events to the committed grant participants. This project also benefited from strong leadership at the district level.

Dissemination Requirements. In another multidistrict grant with widely dispersed sites, participating teachers were encouraged by the project to work in partnership with other participants but also were required to do outreach to nonparticipants. The project provided training on how to conduct outreach, and the grants manager followed up to ensure all participants met this commitment. Grant participants were required to submit plans, document attendance at outreach activities, and submit a final report on the effectiveness of the events. The evaluator estimated that "150 additional teachers were trained or mentored" by grant participants in 2008. Also, in the Annual Performance Report, 12 participants reported being asked by their school or district to conduct a training, and six reported having developed formal mentoring arrangements with other teachers.

As a component of one single-district grant, all of the participating teachers were required to lead or participate in staff development for elementary school teachers, who typically did not have specialized training in social studies. Some of these elementary teachers continued to tap into the knowledge of the participating teachers outside of the staff development. The participating teachers who were interviewed said that they consistently shared with their colleagues whatever materials and resources they were able to bring back from the workshops or the trips.

Teacher outreach and collaboration could evolve into a larger endeavor to build the long-term quality of history teaching at a regional level. The example below illustrates how a TAH grant became the basis for ongoing regional professional development activity.

Use of TAH Funds to Develop a Regional Council of a National Professional Organization

At one of the rural sites, TAH funding was used to establish a regional branch of the Council for the Social Sciences. This group brought together a number of local social science councils, including three rural councils, covering a large area of the state. A central executive committee was formed representing three local areas, each of which had a vice president and smaller boards, who ran their own professional development programs at the local level. This organization was exceptional among the case study sites in that it allowed for greater teacher involvement in the management of their own professional development, leading to more leadership, communication and collaboration among teachers. The council organized an annual conference that has now been held for three years and averages between 150 and 200 participants. The president of the council noted:

"One of the very important parts of the grant was to maintain something—some cohesion, some camaraderie, long-term learning—to enable us to live after the grant... So they put effort and funds into getting a local social studies council going...[so that] the communication and cooperative learning would continue even after the grant was done."

Teacher Recruitment: A Continuing Challenge

Each of the case study grantees reported at least some difficulty recruiting American history teachers who were most in need of professional development. This finding was consistent with findings of the 2005 implementation study of the TAH program. Most case study project directors reported that participants tended to have more years of experience and held more advanced degrees in history than the average American history teacher. At least one grantee reported that very experienced teachers (25 years of experience or more), as well as novice teachers (fewer than three years of teaching) were less likely to participate than those in-between those extremes.

All case study TAH grantees made participation in TAH projects voluntary and used a variety of approaches to recruit teachers. The recruitment process could be lengthy and required a considerable time investment by project staff. This was especially the case for large, multi-district grantees that sometimes encompassed large geographical distances. Project leaders of such grants—often based in county-level education offices—did extensive outreach through contact with superintendents, presentations for teachers and principals at school or district meetings, and invitations to special events at which the project was presented and discussed.

TAH programs asked for a significant commitment of teacher time during and after school hours, on Saturdays, and even during the summer. Highly motivated and engaged teachers who were interested in participating sometimes had multiple prior commitments, such as coaching and other extracurricular activities. But novice teachers, struggling to adapt to teaching and often required to participate in induction programs, were particularly pressed for time. Respondents also cited a reluctance among some more experienced teachers to innovate and try new approaches or new content. As one project leader noted, "We're asking them [the teachers] to go outside their comfort zone," which was difficult for many teachers.

Direct Versus Indirect Recruitment. Grantees often relied on district leaders or principals to communicate with teachers about the grant. However, some principals were reluctant to release their teachers to attend TAH activities and did little to publicize the program among teachers or delayed notifying teachers until after project start-up. Some grantees recruited teachers through the distribution of fliers in faculty mailboxes, emails, presentations at school meetings, or speaking with department chairs and teachers in-person to promote interest in the program. In-person recruitment or recruitment through current or prior participants happy with the program, were among strategies noted by project directors to be successful.

Widening the Pool of Participants. To attract more participants, several programs expanded enrollment to include a wider range of teachers, at additional grade levels or from more widespread districts. At least one grantee used videoconferencing technology to connect the more far-flung districts. Project staff found that it was necessary to accommodate teachers' busy schedules with flexible approaches. Most projects offered duplicative sessions on the same topic so that teachers could choose dates and times that best fit their schedules. Some projects offered different levels of participation; while core participants were required to commit to 40 or more hours of professional development, others teachers were invited to attend single events such as the summer institute or special lectures.

Recruitment Incentives. A few grantees rewarded teachers for participation with laptops and financial incentives. In addition, many of the grantees, particularly those in rural areas with fewer local historical resources, offered a long distance field trip as part of an effort to recruit and retain teachers. Seven of the 16 case study grantees included an out-of-state field trip as part of their programs.

Offering participation incentives—out-of-state field trips was the most frequently cited example of this—undoubtedly contributed to driving up the cost per participant. Analysis of the cost per teacher in TAH projects suggests that field trips can raise the expenditure level to over $30,000 per teacher over three years of participation.[9] This high per participant cost led some of the

---

[9] The cost per participant, based on total number of participants reported in interviews and APRs, varied widely from a low of just over $3,000 to a high of over $10,000 per year based on project expenditures of Year 2 of the grant (2007–08).

respondents to question whether the grant monies could have been used for other purposes with a more direct impact on teaching practices and student performance.

Interviews with teachers and project directors did suggest that field trips to historic sites, including those requiring long-distance travel, provided intensive immersion in American history and were a highly valued component of some TAH projects. Some teachers in western, remote locations reported that first-time trips to Washington, D.C., had a positive impact on their teaching.

Conclusions

While it was not possible to establish clear associations between specific practices and outcomes, the case studies revealed ways in which TAH projects made use of partnerships to enrich the teaching of American history. The case studies also identified teacher recruitment as a major challenge. Even in projects implementing high-quality professional development, the impacts of the projects could be severely limited if the projects reached only more experienced or more innovative teachers.

Chapter 5
Conclusions and Implications

The TAH Program was highly valued by participants at the case study sites. Teachers reported that exposure to the expertise of professional historians and master teachers had increased their knowledge of American history and their historical thinking skills. They often commented that their improved teaching, in turn, had improved student performance and appreciation of history. Many observed that the informal networks of teachers and relationships with universities and history-related professional organizations established by the TAH projects are likely to continue beyond the life of the projects. In some cases, district officials also went out of their way to express their appreciation for the much-needed professional development of their American history teachers.

However, the question of whether the TAH program has an impact on student achievement or teacher knowledge remains unanswered. This study examined TAH outcomes analysis options using extant data: state assessment data and grantee evaluation reports. The study found that a small number of states regularly administer student American history assessments; many states do not have the resources to administer statewide student assessments in subject areas beyond mathematics, reading, and science. TAH grantees are developing new forms of assessment, but these are in the early stages. Furthermore, most TAH grantee evaluations lack rigorous designs. Overall, the data available to measure TAH effects are limited.

Case studies produced suggestive evidence that TAH projects have incorporated a number of practices that have been identified as promising in professional development research. Project directors and participants reported that strong partnerships with organizations rich in historical resources and expertise led to a valuable professional development experience for teachers. Most projects offered a mix of professional development experiences, and some built active teacher learning communities and dissemination networks.

Case study research identified several promising TAH professional development practices that combine history content and pedagogy. Many of these were grounded in an effort to help teachers conceptualize history not as a static unchanging body of knowledge but as an interpretive discipline in which historical analysis and interpretation can result in multiple perspectives on historical events. By modeling approaches for using primary source documents in the classroom (such as through think-aloud protocols, questioning strategies, and the use of multiple documents with differing perspectives), master teachers were able to demonstrate how much reliance on a textbook limits options for teaching history. Several practices such as lesson plan development using primary sources, original teacher research, and project-based instruction in which students uncover local history through primary sources, all helped teachers obtain a deeper understanding of the work of historians and communicate this to students.

Case studies also revealed areas in which projects were struggling. Projects continued to face the challenge of recruiting teachers most in need of support. While a benefit of TAH programs is that they offer an alternative to the single session workshop model, the extensive commitment of time and effort required by many projects meant it was often difficult to fill all available slots for participants. Some projects recruited teachers by offering extensive field trips to out-of-state historical sites. While teachers benefited from the visits, the cost per participant was sometimes

excessive. An additional recruitment approach was to offer teachers a tiered menu of offerings that allowed for varying levels of time and commitment. In those cases, teachers were able to select a level of participation that matched their personal circumstances.

Lack of active support or involvement of school or district leaders was another challenge facing many case study projects. Strong support by district or school leaders in a few projects eased the process of recruitment, dissemination, and integration of the project with other district activities and priorities. More typically, such support was weak or lacking. In a few cases participants faced difficulties obtaining approval for release time for TAH professional development.

All of these key findings have important implications for the TAH program in the future. Clearly, the characteristics of strong projects could be incorporated into future projects' planning, development, and proposal processes, as well as the Department's criteria for awarding and monitoring grants. In addition, the research highlights two particularly stubborn challenges for the TAH program since its inception: (a) measuring the impact of the program and (b) recruiting the teachers most in need of improving their skills and knowledge.

Measuring Impact

As the TAH projects have shifted toward a greater emphasis on skills in historical analysis and inquiry, state American history assessments may be less appropriate as outcome measures for the program. Moreover, many states do not have American history assessments, are engaged in revising them, or have suspended their administration as a cost-cutting measure. The resulting mismatch between available standardized assessments and the work of the TAH projects makes it difficult for local or national evaluators to measure project outcomes accurately.

As this evaluation shows, teacher and student outcome measures remain elusive. Some case study grantees and their evaluators have developed project-based assessments that measure both historical thinking skills and content knowledge. However, many lack the funding, time, and expertise to further refine, pilot, and validate those assessments and to find cost-effective approaches to administering and scoring such tests.

Federal investments could be useful in several ways. First, an investment could be made in bringing together evaluators with first-hand experience in developing innovative assessments, along with other assessment experts, so that existing expertise could be shared and extended. Second, investments in the further development, validation, and dissemination of models for teacher and student assessment tools that could be shared across projects could contribute both to stronger local evaluations and to potential comparisons between projects. Submission of electronic lesson plans or assessment forms to central sites for scoring could potentially reduce costs for individual grantees. In addition, a more standardized approach to tracking project participation and to linking students' outcome data to teachers, would support cross-site outcomes analysis. More rigorous evidence of the impact of the various TAH program models could then be generated. Currently, the lack of a common approach to reporting the yearly number of participants and their total hours of participation has limited efforts to collect data on the relationship between duration of participation and other outcomes.

Even with better measurement tools, local evaluators are likely to struggle with the identification and recruitment of appropriate comparison groups. For local evaluations to be successful, comparison groups must be built into the design of the projects. Thus, awarding grants based on

the strength of the applicants' research designs is more likely to result in solid measures of grantee outcomes.

Strengthening Recruitment and Participation

Although teachers with a variety of backgrounds have participated in the TAH programs, TAH projects have struggled with recruiting American history teachers who are most in need of improvement. Given the serious commitment of time and energy required of participating teachers, fuller integration of the program into schools or districts may be necessary in order to reach teachers at all levels. School-based approaches, which were rare among the case study programs, could reduce the amount of time for professional development outside of the regular school day and contribute to sustained reform. Application priorities in recent years have targeted grants to schools identified for academic improvement. Schoolwide approaches would particularly benefit these schools.

Participants in the TAH program have reported that the professional development it offers is of high quality, is useful in the classroom, and enables them to engage students in an improved understanding of history and historical inquiry. The program could be improved with new approaches to teacher recruitment and to schoolwide or districtwide commitment. Assessment of the impact of TAH may be possible with increased evaluation rigor and further development or validation of student learning measures in American history.

References

Almond, D. and J. Doyle Jr. 2008. *After midnight: A regression discontinuity design in length of postpartum hospital Stays.* NBER Working Paper.

American Council of Trustees and Alumni. 2000. *Losing America's memory: Historical illiteracy in the 21st century.*
http://www.goacta.org/publications/Reports/acta_american_memory.pdf (accessed June 18, 2003)

Anderson, S.E. 2003. *The school district role in educational change: A review of the literature.* Toronto, Ontario: Ontario Institute for Educational Change, ICEC Working Paper #2.

Apt-Perkins, D. 2009. Finding common ground: Conditions for effective collaboration between education and history faculty in teacher professional development. In *The Teaching American History Project: Lessons for history educators and historians.* Ed. R.G Ragland and K.A. Woestman. New York: Routledge.

Bain, R. 2005. They thought the world was flat: Applying principals of how people learn in teaching high school history. In *How students learn history in the classroom.* Ed. National Research Council. Washington, DC: National Academies Press.

Bell, C., L. Brandon, and M.W. Weinhold. 2007. New directions: The evolution of professional development directions. *School-University Partnerships: The Journal of the National Association for Professional Development Schools,* 1 (1) 45–49.

Bennis, W. and B. Nanus. 1985. *Leaders: The strategies for taking charge.* New York: Harper and Row.

Berkeley Policy Associates. 2005. Study of the Implementation of Rigorous Evaluations by Teaching American History Grantees. Oakland, CA: Berkeley Policy Associates. Unpublished manuscript.

Berkeley Policy Associates. 2007. *Teaching American History Evaluation, Technical Proposal.* Submitted to: U.S. Department of Education. Oakland, CA: Berkeley Policy Associates. Unpublished manuscript.

Berkeley Policy Associates. 2008. *Feasibility Study of State Data Analysis: Teaching American History Evaluation.* Submitted to: U.S. Department of Education. Oakland, CA: Berkeley Policy Associates. Unpublished manuscript.

Bloom, B.S., Ed. 1956. *A taxonomy of educational objectives: The classification of educational goals.* Susan Fauer Company, Inc.

Bloom, H. 2009. "Modern Regression Discontinuity Analysis." MDRC Working Papers on Research Methodology, New York: MDRC.

Bohan, C. H., and O.L. Davis, Jr. 1998. Historical constructions: How social studies student teachers' historical thinking is reflected in their writing of history. *Theory and Research in Social Education*, 26, 173–197.

Carver, C.L. 2008. Forging high school-university partnerships: Breaking through the physical and intellectual isolation. *School-University Partnerships: The Journal of the National Association for Professional Development Schools.*

Coburn, C.E. 2004. Beyond decoupling: Rethinking the relationship between the instructional environment and the classroom. *Sociology of Education,* 77 (3), 211–244.

Cruse, J. M. 1994. Practicing history: A high school teacher's reflections. *The Journal of American History,* 81, 1064–1074.

Darling-Hammond, L. and M. McLaughlin. 1995. Policies that support professional development in an era of reform. *Phi Delta Kappan,* 76 (8), 597-604.

Desimone, L.M., M.S. Garet, B. F. Birman, A. Porter, and K.S.Yoon. 2003. Improving Teacher In-Service Professional Development in Mathematics and Science: The Role of Postsecondary Institutions. *Educational Policy* 17 613-648.

Desimone, L.M., A.C. Porter, M.S. Garet, K.S.Yoon, and B.F. Birman. 2002. Effects of Professional Development on Teachers' Instruction: Results from a Three-year Longitudinal Study. *Educational Evaluation and Policy Analysis* 24 (Summer): 81–112.

Duttweiler, P.C. and S.M. Hord. 1987. *Dimensions of effective leadership*. Austin, TX: Southwest Educational Development Laboratory.

Education Week. 2006. *Quality counts at 10: A decade of standard-based education*. Editorial Projects in Education Research Center.

Elmore, R.F. 2004. School Reform from the Inside Out: Policy, Practice, and Performance. Cambridge, MA: Harvard Education Press.

Garet, M.S., A.C. Porter, L.M. Desimone, B.F. Birman, and K.S. Yoon. 2001. What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal* 38 (Winter): 915–945.

Glass, G. V., B. McGraw, and M.L.Smith. 1981. *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Gess-Newsome, J. and N.G. Lederman. 1999. *Examining pedagogical content knowledge.* Boston: Dordrecht: Kluwer Academic Publishers.

Gleason, P., M. Clark, C.C. Tuttle, and E. Dwoyer, 2010. *The evaluation of charter school impacts: Final report.* U.S. Department of Education, Institute of Education Sciences, Washington, D.C..

Grant, S. G. 2001.It's just the facts, or is it? The relationship between teachers' practices and students' understandings of history. *Theory and Research in Social Education*, 29, 65–108.

Hamilton, L. M., B.M. Stecher, and S.P. Klein. 2002. *Making sense of test-based accountability in education.* Santa Monica, CA: Rand. *http://www.rand.org/publications/MR/MR1554/* (accessed June 3, 2003)

Hartzler-Miller, C. 2001. Making sense of "best practice" in teaching history. *Theory and Research in Social Education,* 29(4), 672–695.

Hassel, E. 1999. *Professional development:learning from the best*. Hassel, E. Oak Brook, IL: North Central Regional Educational Laboratory.

Hooper-Greenhill, E. 2004. *The Educational Role of the Museum*. New York: Routledge.

Hunter, J.E., and F.L. Schmidt. 1990. Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, *75*, 334–349.

Jackson, R., A. McCoy, C. Pistornio, A. Wilkinson, J. Burghardt, M. Clark, C. Ross, P. Schochet, and P. Swank. 2007. *National evaluation of Early Reading First: Final report.* U.S. Department of Education.

Kobrin, D., S. Faulkner, S. Lai, and L. Nally. 2003. Benchmarks for high school history: Why even good textbooks and good standardized tests aren't enough. *AHA Perspectives* 41 (1). *http://www.historians.org/perspectives/issues/2003/0301/0301tea1.cfm* (accessed April 7, 2010).

Knupfer, P.B. 2009. Professional development for history teachers as professional development for historians. In *The Teaching American History Project: Lessons for history educators and historians,* ed. R.G. Ragland, and K.A. Woestman. New York: Routledge.

Kubitskey, B., and B.J. Fishman. 2006. A role for professional development insustainability: Linking the written curriculum to enactment. In *Proceedings of the 7th International Conference of the Learning Sciences,* Vol. 1, ed. S. A. Barab, K. E.Hay, and D. T. Hickey, 363–369. Mahwah, NJ: Lawrence Erlbaum.

Lancaster, J. 1994. The public private scholarly teaching historian. *The Journal of American History,* 81(3), 1055–1063.

Lee, J., and A. Weiss. 2007. *The Nation's Report Card: U.S. History 2006* (NCES 2007-474). U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Leming, J., L. Ellington, and K. Porter. 2003. *Where did social studies go wrong?* Washington, D.C.: Thomas B. Fordham Foundation.

Liberman, A., and M.W. McLaughlin. 1992. Networks for educational change: Powerful and problematic. *Phi Delta Kappan* 73: 673–677.

Lipsey, M.W., and D.B. Wilson. 2001. *Practical meta-analysis*. Thousand Oaks, CA: Sage.

McLaughlin, M., and J.E. Talbert. 2006. Building school based teacher learning communities: Professional strategies to improve student achievement. New York: Teachers College Press.

National Center for Education Statistics. 1996. *Results from the NAEP American history assessment—At a glance.* Washington, DC: *http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=96869* (accessed Sept. 17, 2003).

National Center for Education Statistics. 2002. *American history highlights 2001 (The Nation's Report Card).*Washington, DC: *http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002482* (accessed Sept. 17, 2003).

National Center for Education Statistics. 2007. *The Nation's Report Card: American History 2006.* Washington, DC: *http://nces.ed.gov/nationsreportcard/pubs/main2006/2007474.asp* (accessed April 17, 2010).

Newman, F.M., A.S. Bryk, and J. Nagaoka, J. 2001. *Authentic intellectual work and standardized tests: Conflict or coexistence.* Chicago, IL: Consortium on Chicago School Research. University of Chicago.

Newman, F.M. and Associates.1996. Authentic achievement: Restructuring schools for intellectual quality. San Francisco: Jossey-Bass.

Newman, R.S. and G.H. Wehlage. 1995. Successful school restructuring: A report to the public and educators by the Center on Organization and Restructuring of Schools. Alexandria, VA: Association for Supervision and Curriculum Development.

Paige, R. 2002, May 9. Remarks on NAEP history scores announcement. Retrieved Aug. 26, 2002, from: http://www.ed.gov/Speeches/05-2002/05092002.html.

Penuel, W.R., B.J. Fishman, R.Yamaguchi, and L. P. Gallagher. 2007. What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal* 44 (December): 921–958.

Penuel, W.R., K.A. Frank, and A. Krause. 2006. The distribution of resources and expertise and the implementation of schoolwide reform initiatives. In *Proceedings of the Seventh International Conference of the Learning Sciences. Vol. 1,* eds. S.A. Barab, K.E. Hay and D.T. Hickey,522–528. Mahwah, NJ: Lawrence Erlbaum.

Peterson, G. 1999. Demonstrated actions of instructional leaders: An examination of five California superintendents." *Education Policy Analysis Archives* 7, no. 18, *http://epaa.asu.edu/ojs/article/viewFile/553/676*. (accessed April 7, 2010).

Raudenbush, S.W., and A.S. Bryk. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage.

Ravitch, D. Aug. 10, 1998, August 10. Lesson plan for teachers. Washington, DC: *Washington Post*: http://www.edexcellence.net/library/edmajor.html (accessed Aug. 23, 2002)

Ravitch, D. 2000. The Educational Background of History Teachers. In *Knowing, Teaching and learning history: National and International Perspectives,* ed. P.N. Stearn, P. Seixas, and S. Wineburg,143–155. New York: New York University Press.

Resnick, L. B., and T.K. Glennan. 2002. Leadership for learning: A theory of action for urban school districts," In ed. A. T. Hightower, M. S. Knapp, J. A. Marsh and M. W. McLaughlin *School districts and instructional renewal.* New York: Teachers College Press.

Sass, T., and D. Harris. 2005. Assessing Teacher Effectiveness: How Can We Predict Who Will Be a High Quality Teacher. Gainesville, FL: Florida State University.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J.R., Porter, J., Smith, J. 2010. Standards for Regression Discontinuity Designs. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.

Seixas, P. 1998. Student teachers thinking historically. *Theory and Research in Social Education,* 26, 310–341.

Shulman, L.S. 1987. Knowledge and teaching: Foundations of the new reform. *Harvard Education Review,* 10 (1), 9–15, 43–44.

Slekar, T.D. 1998. Epistemological entanglements: Preservice elementary school teachers' "apprenticeship of observation" and the teaching of history. *Theory and Research in Social Education*, 26, 485–508.

Smith, J., and R.G. Niemi. 2001. Learning history in school: The impact of course work and instructional practices on achievement. *Theory and Research in Social Education,* 29, 18–42.

Snipes, J., F. Doolittle, and C. Herlihy. 2002. *Executive summary. Foundations for success:* Case studies of how urban school systems improve student achievement. New York: MDRC.

St. John, M., K. Ramage, and L. Stokes. 1999. A vision for the teaching of history-social science: Lessons from the California History-Social Science Project. Inverness, CA: Inverness Research Associates.

Steans, P., and N. Frankel. 2003. Benchmarks for professional development in teaching of history as a discipline. *Perspectives Online* 41, no. 5. *http://www.historians.org/perspectives/issues/2003/0305/index.cfm* (accessed April 7, 2010).

Stearns, P.M., P. Seixas and S. Wineburg. 2000. *Knowing, teaching and learning history: National and international perspectives*. New York: New York University Press.

Teitel, L. 1994 Can school-university partnerships lead to the simultaneous renewal of schools and teacher education? *Journal of Teacher Education* 45: 245–52.

Thornton, S.J. 2001. Subject specific teaching methods: History. In ed. J. Brophy. *Subject specific instructional methods and activities,* 229–314. Oxford, U.K.: Elsevier.

Tichenor, M., M. Lovell, J. Haugaard and C. Hutchinson. 2008. Going back to school: Connecting university faculty with K–12 classrooms. *School-university Partnerships: The Journal of the National Association for Professional Development Schools.*

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. 2005. *Evaluation of Teaching American History program.* Washington, D.C.

Van Hoever, S. 2008. The professional development of social studies teachers. In ed. L. Levstik and C. Tyson, *Handbook of research in social studies education,* 352–372. New York: Routledge.

Van Sledright, B.A. 2004. What does it mean to think historically…and how do you teach it? *Social Education,* 68(3), 230–233.

Watson, N.H. and M.G. Fullan. 1991. Beyond school-university partnerships: In eds. M.G. Fullan, A. Hargreaves, *Teacher development and educational change.* Lewes, DE: Falmer.

Wineburg, S. 2001. Historical thinking and other unnatural acts: Charting the future of teaching the past. Philadelphia: Temple University Press.

Wilson, S. 2001.Research on history teaching. In ed. V. Richardson, *Handbook of research on teaching* (4th ed, 527–544).Washington, D.C.: American Educational Research Association

Woestman, K.A. 2009. Teachers as historians: A historian's experience with TAH projects. In *The Teaching American History Project: Lessons for history educators and historians.* Ed. R.G. Ragland and K.A. Woestman. New York: Routledge.

Appendix A

Case Study Site Selection and Site Characteristics

Case Study Selection

A total of 16 grantees from the 2006 cohort were selected for case study research. Eight grantees were selected to focus on the question: "What are the associations between TAH practices and changes in student achievement in American history?" This selection was based on student American history assessment data provided by the five states also providing data for the state data analysis: California, Texas, New York, Virginia, and Georgia. Grantees were compared by calculating the differences in the z-scores of the mean assessment scaled scores of participating schools between 2005 and 2008.[10] Z-scores measure the difference in the score from the sample mean in standard deviation units and allow us to standardize mean assessment scores across states.

Grantee districts were grouped according to the following categories:

- Previously high-achieving districts that experienced a large change in assessment scores (category=1).

- Previously low-achieving districts that experienced a large change in assessment scores (category=2).

- Previously high-achieving districts that experienced no change or decline in assessment scores (category=3).

- Previously low-achieving districts that experience no change or decline in assessment scores (category=4).

Through pairing case study grantees from categories 1 and 3 and from categories 2 and 4 within the same states, it was possible to compare grantees with improvements in test scores to those with no improvement, while controlling for pre-TAH differences in test score performance. This approach also controlled for contextual and socioeconomic variables to some extent. Grantees with lower preprogram scores (categories 2 and 4) tend to have higher poverty rates than those with higher preprogram scores (categories 1 and 3). By matching grantees based on their preprogram scores as well as their state or region,[11] the pairing made it possible to focus on differences in grantee practices that might influence post-program student test scores.

Eight grantees were selected to focus on the question: "What are the associations between grantee practices and gains in teacher knowledge?" To select these grantees, the study team reviewed 2008 APRs. A total of 119 APRs of the 2006 cohort of TAH grantees were reviewed. During an initial round of reviews, each APR was coded to identify evaluation designs, types of teacher assessments, types of analyses, and findings reported.

Based on the coding and additional review of selected documents, grantees were initially identified that met the following criteria:

---

[10] The lead district from each of the identified 2006 grantees was analyzed. The following grantees were excluded: three Texas grantees, each of which encompassed large numbers of districts (approximately 13), and one grantee in California implemented in a single school, due to non-comparability with other grantees—which generally include between one and three districts.

[11] One pair of case studies could not be matched by state.

- Grantees reported gains in teacher content knowledge, supported by data.

- Grantees reported administration of a teacher content knowledge assessment that was based primarily on items from a national or state standardized test, such as the Advanced Placement Exam, the NAEP, the SAT, the New York Regents Exam, or the California Standards Test.

- Score improvements were reported for participants based on a quasi-experimental evaluation design. Although most evaluations relied on a single-group pre-post design, a small number (three) used comparison groups with some statistical controls.

- Results suggested participation in the TAH program was associated with teacher knowledge gain, although a causal relationship could not be inferred.

Four grantees in four different states were selected that met the above criteria. Each of these sites was "matched" with another site within the same state that had similar demographic characteristics and did not provide evidence of teacher knowledge gains on its 2008 Annual Performance Report.

There were several limitations and biases inherent in the selection process used to identify case study grantees. Factors other than the TAH program might have been responsible for changes in students' American history scores during the 2005–08 period. Evidence of changes in teacher knowledge was based on grantees' self-reported outcome data. It was not feasible to review the teacher tests according to the level of difficulty of test items or how reliably the data were scored and reported. The amount of data regarding the test content, design and reporting varied by annual performance report. In addition, both student and teacher outcomes data were necessarily limited to 2008 data, and therefore reflected only two years of program performance. Because the grants are three years in duration, more complete outcomes data would have been available for 2006 grantees at the conclusion of the grant period in 2009. However, as mentioned earlier, site visits could take place only while the programs were still in operation.

Site Characteristics

The selection process described above resulted in eight pairs of grantees; all but one pair were matched by state. Each pair included one "higher performing" and one "typically performing" site, as identified in the outcomes analysis described above. Identical site protocols were used at all sites. Exhibit 3 presents several characteristics of the sites.

Exhibit 3: Case Study Site Characteristics

Pair	State	Number of Districts	Rural/ Urban	Grade Levels Served	State History Test	Summer Institute
Pairs selected based on student outcomes						
1	New York	1	Urban	MS, HS	Yes	2 week
1	New York	1	Urban	All	Yes	1 week
2	Texas	15	Rural	5-8,10,11	Yes	1 week
2	Texas	1	Rural	MS, all	Yes	None
3	California	17	Rural	All	Yes	2 week
3	California	1	Suburban	4,5,8,11	Yes	None
4	New York	68	Rural, Suburban	4-8; 11, 12	Yes	None
4	California	14	Urban, Rural	MS, HS	Yes	None Symposium
Pairs selected based on teacher outcomes						
5	Maryland	1	Urban, Suburban	MS, HS	No	2 week
5	Maryland	3	Rural, Urban	MS, HS	No	1 week
6	Kentucky	1	Urban	HS	Yes	1 week
6	Kentucky	14	Rural	5,8	Yes	1 week
7	Ohio	35	Urban, Rural Suburban	All	No, undergoing change	2 week
7	Ohio	8	Urban, Suburban	All, HS	No, undergoing change	2 week
8	Massachusetts	3	Urban, Suburban	All, mostly HS	No, discontinued	2 week
8	Massachusetts	4	Small Urban	All	No, discontinued	1 week

Appendix B

Evaluation Review

Additional Technical Notes and Exhibits

Exhibit 4: **Summary Description of 94 Evaluation Reports Reviewed in Stage I**

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
1	1	Analysis of state social studies test scores for matched cohorts of students of teachers who participated in TAH and students of non-TAH teachers for third- through eighth-graders. Tests of significance, mixed model framework for analysis with fixed and random effect controls, etc. *No discussion of the fact that this is a social studies test; not specifically an American history test; no alignment of test with treatment content.* No final evaluation report. APR data only.	Pre, Post 1, and Post 2 teacher content test, 23 multiple-choice and 2 constructed response items, drawn from NAEP or state assessment items. Each test included different content, matching what was covered in the most recent PD.
2	0	Pre-post only. No control group. SAT 10 for Grade 6 and Alabama High School Graduation Examination for Grades 10 and 11. History section of the SAT 10 analyzed separately for subsample. Very limited data in APR.	Multiple-choice items from AP College Board and NAEP used as pre-test and post-test for teachers. No details.
3	0	This is a year 4 report without student performance data included, but a 1–3 year report is mentioned that has a quasi-experimental design.	
4	1	Quasi-experimental study of student achievement using treatment and comparison group design.	
5	1	Year 4 –NY State Regents U.S. History and Government test data analyzed for students of 4 teachers. No citywide data available for this year (2007–08). Data collected in 2005, 2006 and 2007 from participating teachers each year (sample sizes of 300–400) and compared to district outcomes. No evaluation report. Aggregated data only in APR report.	
6	0	No student achievement data analyzed. Teacher knowledge based on 45 item pre-post test.	

Continued on the next page

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
7	1	Treatment and control groups for elementary, middle, and high school students. Test included three measures- disciplinary concepts, construction of knowledge, elaborated written communication. Teacher Assignment/Student work evaluations were conducted based on Newman's work on authentic intellectual achievement. MANOVA on high school sample with academic ability as covariate, Small sample sizes. Evaluation Report has relatively complete description of design and results.	Teachers' elaborated written communication on history topics was also evaluated.
8	0	Very short "extension of project" report. Report alludes to administration of TX state test in Grades 8, 10, and 11 but no details provided.	
9	0	Summary APR report only. Brief reference to previous studies of student achievement on Texas U.S. history test but no data provided.	
10	1	Quasi-experimental analysis of student achievement in history using control and comparison groups of middle and high school students. Throughout the 2007–08 academic year, 545 students from the classrooms of 26 participating teachers and 287 students from classrooms of 13 nonparticipating teachers were administered project developed grade appropriate history tests. Pre- and post-tests were administered. Minimal reporting on design and data results.	Pre- and-post teacher content test was administered. Minimal reporting on design or data results.
11	0	None	
12	0	None	
13	1	Student achievement analysis examines New York State Social Studies Exam results for Grades 5, 8, 11. Compares project students vs. non-project students across district. Good reporting of data.	
14	1	Evaluation conducted by RMC. Full report included.	Evaluation conducted by RMC. Full report included.
15	0	No experimental component to student or teacher performance assessment. Student data was examined across district over time.	

Continued on the next page

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
16	1	Data collected in 2007–08 from 3,118 high school students and 918 eighth-grade students on selected Nebraska American history standards. Comparison of students of treatment and non-treatment teachers. Minority student data analyzed. Research design details limited.	
17	1	Evaluation conducted by RMC with strong quantitative and qualitative data. Full report included.	Evaluation conducted by RMC. Complete report included.
18	0	Longitudinal student achievement analysis included in plan of work, but the report itself includes no analysis. There was no control group.	
19	1	Used NAEP exam test items addressing historical periods covered by the treatment. Test items included: 30 multiple choice, five essays. Treatment and control groups were included. 2004–05 was baseline; data collected in 2006 and 2007 served as a comparison of students matched to teachers. NC end-of-course test being restructured. Baseline data collected 2005-06 and data collected 2006–07. Some reporting of AP test scores. Research design and description of results were limited (e.g. grade levels of students unclear).	Unclear. Did not appear to have teacher content test. Teachers kept portfolios.
20	0	None	
21	0	No standardized assessment given to students because teachers felt it was a "forced fit."	
22	1	Subset of NAEP test administered to students of participants and control group. Data collected in 2006–08. Limited data in APR. Mean scores reported. No reported significance testing	Pre- and post- test, 30 questions with a "mix of 8th- and 12th-grade questions" Treatment and control groups. Description of test and results is unclear.

Continued on the next page

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
23	1	Quasi-experimental study of results of scores on Kentucky Core Content Test in Social Studies (history, economics, geography, interpretation components) with treatment and control. Approx. 625 students in each group. However 2007 version of test is new and uses a different scale that cannot be linked to previous year's performance. Also 30 percent attrition rate of treatment teachers.	Pre-post content knowledge measures included a test of critical thinking, an extended response item test to assess evidence-based interpretations, and a historical thinking survey with short answer responses. Three years of results were presented. Control group data was collected.
24	0	Some pre-post testing in some grades. No controls. Limited data.	Pre- and post- teacher content test. No description.
25	0	No student achievement analysis conducted.	
26	1	Quasi-experimental design was used for student achievement analysis; A pre-post assessment with control group (only five teachers participated) was implemented.	
27	0	Study design was weak; quasi-experimental design (program student achievement vs. nonprogram student achievement, snapshot) and not using a standardized (or specified) measure of student performance.	
28	0	None	Self-report surveys only.
29	0	None. Data on classroom observations, course grades, and student failure rates compared with controls.	Self-report and classroom observations only.
30	1	Released items from Massachusetts, Texas, and New York Regents exams aligned to Kentucky's standards for elementary and middle school. Administered in Fall and Spring, 2004–05, 2005–06, 2006–07, and 2007–08 (new student cohorts each year). Alpha reliability estimates for the tests were reported. *Kentucky Core Content Test* (KCCT) – school scores on Social Studies portion were also collected and compared to district and state averages.	Not mentioned in extension report.
31	0	None.	None. Survey only.

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
32	0	Pre-post only. No controls but strong evaluation report with multiple qualitative test results. Complete copies of tests, surveys, scoring rubrics.	
33	0	None or minimal. There is a Kentucky Core Content Test for Social Studies. Minimal reporting of 2007 test results at the school level. Test significantly restructured in 2007.	Pre-post summer institute content test 2004–07 designed by project professors. No controls. Limited description.
34	0	Eleventh grade State Subject Area Testing scores in U.S. History since 1877 reported and compared to state percentages and to scores from previous years (2005, 2006 and 2007 exit exam). Not enough data to evaluate design or findings.	Extensive teacher survey self-reports.
35	0	This Year 4 report mentions that the Year 3 report described student achievement data for middle and high school based on a district assessment that included an essay. Description is vague. The n may have been small (e.g. one participating and one control classroom).	University professor developed pre- and post- test related to summer graduate course. Control group teachers also took post-test. Lesson plan evaluations (with rubric), surveys and other measures compared to randomly selected control group with significance testing.
36	1	Treatment and comparison groups of the 2006–07 cohort were administered a pre-post test. Grades 5, 8, and 11 were included. (The 2008 cohort was not tested.) Tests of significance were performed. Note that test items (apparently for all grades) were developed based on questions submitted by district AP teacher. 2008 exit level test data for participating and non-participating school districts were compared. (Limited details.)	Locally developed pre-post content based on AP test.
37	0	No data.	
38	0	No student assessment data reported. Self-report surveys only.	Questionnaires and surveys only.

Continued on the next page

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
39	0	No quasi-experimental design. Very inconsistent student data reporting. The type of tests changed in the middle of the grant period.	
40	0	2004-07 pre- and post-test based on ACT American history assessment items (multiple choice). Assessment was administered in grades 4–12; however minimal or no data reported. No control groups.	Pre- and post- teacher test based on 40 ACT assessment items for American history. 2004–07 administered along with attitude survey. Minimal data reported.
41	0	No quasi-experimental design for student achievement analysis was included in their proposal or their report.	Quasi-experimental design for assessing teacher content knowledge,
42	1	Eighth- and 11th-grade students of TAH and non-TAH teachers within one district compared on CST history and ELA tests in Years 2, 3 and 4 (2007–08). Tests of significance performed. DBQs (document-based questions to measure historical thinking skills) administered in fall and spring of 2007–08 to Grade 11 students; scored by external evaluators. Data tables available with test results and scores.	California Teachers of American History Content Assessment given regularly. (This is a teacher self-report survey used in several CA projects with projects presenting "quantitative data" based on self-assessment of change.) Locally developed multiple-choice assessment of teacher content knowledge developed by CAL State professors in 2007. Twenty-five questions, pre- and post- summer institute.
43	0	No experimental or quasi-experimental design. No state American history test in Oregon reported.	Teacher self-report survey only.
44	0	No student performance data included in report	
45	0	No experimental or quasi-experimental design. Review of random sample of student work using rubric, with scores reported and classroom observation with scores reported. Student self-report survey.	Self-report surveys with scores derived.

Continued on the next page

Continue from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
46	0	APR reported on last year no-cost extension only; incomplete data. No mention of student achievement data.	Modified version of AP History exam in use with limited teacher gains; however no scores or details mentioned.
47	0	No student achievement analysis included. Extensive survey results on summer workshop	
48	0	Used longitudinal TAKS scores across the district as a proxy for program evaluation.	
49	0	Grades 5, 8, 11 tests using NY State Elementary Test of Social Studies and N.Y. State Regents U. S. History and Government tests. Participating district and school scores compared with nonparticipating districts. Fifth- and eighth-grade data collected on level change, no individual scores. Chi-square analysis was conducted. At Grade 11 there were 416 students, one TAH school and one non-TAH school compared.	Teacher content knowledge test mentioned as one part of teacher portfolio of outcomes but no description.
50	1	Student data collected 2005–07 (not 2008) in CST English and History or Soc. St. grades 8–11. 2007 data reported in APR. Project students compared to students with non-TAH teachers. Scale scores reported on tests and subtests. Analysis of variance conducted and reported, tests of significance, mean, standard deviation, and errors. Significant positive results. Conducted history writing assessment 8th and 11th grade twice per year 2005–07 (with pilot in 2004). Comparison groups in spring 2005 and 2006 with TAH students scoring significantly higher.	
51	0	Little information provided.	

Continued on the next page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
52	0	No experimental or quasi-experimental done.	
53	0	No student assessment data. All performance results based on teacher self-survey.	All performance results based on teacher self-survey.
54	1	Good reporting of data. Student performance based on piloted measure using NAEP items. A project evaluation report is attached to the final performance report.	
55	0	No experimental or quasi-experimental design used. Districtwide changes in student achievement are the measures used to proxy student performance.	
56	0	Extremely limited reporting with almost no explanation of measures or sampling structure.	
57	0	Student achievement results based on schoolwide data. No further detail.	
58	0	No experimental component. Student achievement analysis based the change in district scores over time.	
59	0	No student assessment information.	
60	0	Student achievement analysis based on comparing schoolwide data (including teachers who participated and those who didn't) over time. Not an experiment.	
61	1	Quasi-experimental design with comprehensive suite of assessment measures in grades. CST is the standardized test.	

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
62	0	No student achievement data recorded because of the small district (most teachers were the only teachers at that grade level).	
63	0	Strong qualitative evaluation done by a third-party evaluator. No reported quantitative results.	
64	1	Clear reporting of data. Student achievement results based on comparison of participating teachers' students compared to non-participating teachers' students across several districts.	Teacher content knowledge assessments reported as "in progress."
65	0	No student achievement data collected.	Teacher content test: 40 TAH teachers, New Hampshire teachers as control. Highly limited description of teacher test design or results.
66	0	No student assessment taken in this no-cost extension year.	
67	0	Very unclear reporting of results. Use of an experimental design but no quantitative data reported, including no information about student assessment.	
68	0	State American history assessment was discontinued. Used a project created measure with entire district as a control. The assessment results were not analyzed using a quasi-experimental design. No quasi-experimental assessment of project results.	
69	0	Student knowledge assessment based on nonstandardized pre-post test.	No measure of teacher content knowledge
70	0	No assessment of student knowledge,	Results based on teacher self reports and attitudes.
71	1	Student achievement measured by pre-post content tests and school and district level data.	Teacher content knowledge assessed based on pre-post test scores.
72	1	Student achievement measured with large (2000) comparison group on a measure that was based on state standards.	

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
73	1	Quasi-experimental student achievement analysis based on project created assessment. Control and experimental groups matched on "demographic similarity."	
74	0	No performance data for students.	Teacher results based on self-survey.
75	0	Minimal reporting.	
76	0	Student achievement based on results from two participating teachers' classrooms. Teachers were selected for the program for their leadership qualities. Presence of sampling bias. Student achievement analysis done comparing schoolwide performance to district performance on Regents Exam.	
77	0	No student performance measures included.	
78	1	Student achievement data reported. Use of a quasi-experimental design including matched students and CST tests. Clear reporting of results.	Teacher content knowledge measured by self-assessment.
79	0	No student achievement data included; most of the packet contains curriculum examples.	
80	1	Quasi-experimental design using matched comparison groups.	
81	0	No evaluation of student achievement.	Teacher pre-post assessment based on AP test.
82	1	Quasi-experimental design, students' performance measured against matched controls, clear reporting of findings.	Teachers' efficacy measured via survey.
83	1	Quasi-experimental design, students assessed in TAH teachers classes before and after TAH training (different students each year).	Teachers assessed pre-post training based on selected AP questions.
84	0	No quasi-experimental or experimental design.	

Continued from the previous page

Grantee	Rigorous? (0 = no, 1 = yes)	Study Design and Student Achievement Outcome Data	Teacher Test of American History Content
85	0	Although a quasi-experimental design is discussed in the performance report, the grantee says that the data is not yet available, and that they will send the experimental data in a future report.	
86	0	No student achievement data collected.	
87	1	Student achievement results were in a quasi-experimental design, experimental group was students in TAH teacher classrooms, control was students of another teacher at the same school (one control teacher for each experimental teacher). Results based on project-designed test.	Teacher content knowledge results based on self-survey.
88	0	No experimental or quasi-experimental design for students.	Teacher content knowledge assessment based on attitudinal self-survey.
89	1	Student achievement results were in a quasi-experimental design. Control group was district wide performance on state history exam, treatment was just students of TAH participating teachers.	
90	1	Quasi-experimental design with large control and experimental groups (about 1,100 students in each). Test given was described as "standards based and standardized" but it is not specified. Analysis was performed for both student and teacher content knowledge.	Teacher content knowledge results based on self-survey,
91	0	Student achievement results were in a quasi-experimental design; there were only two teachers in the treatment group.	Teacher content knowledge results based on self-survey.
92	1	Control group used for student results, inferential statistics, clear reporting of findings.	
93	0	No quasi-experimental design for student results.	Teacher content knowledge assessed based on pre-post self reporting.
94	1	Quasi-experimental design used, clear reporting of findings.	

Exhibit 5: Summary Description of 32 Evaluation Reports Reviewed in Stage 2

Grantee	Assessment	Comparison Type	N for each group	Mean by group	SD for each group	Effect Size	T/F Statistic	Multiple grades
1	South Carolina Statewide	Treatment vs. Control	Yes	Yes	Yes	No	No	Yes
2	TAKS Texas Statewide Social Studies Test	Treatment vs. Control	Yes	Yes	Yes	No	Yes	Yes
3	NY State Regents U.S. History and Government test	Treatment vs. Constructed Comparison group at the district level	Yes	No - mean percent, not mean score	No	No	No	Yes
4	Student Work Newman and Bryk	Low-PD vs. High-PD - risk of self selection bias	No	Yes	No	Yes	Yes	Yes
5	Project Developed + Reading and Writing on CT Statewide Test	Treatment vs. Control	Yes	Yes	No	No	Yes	No
6	New York State Social Studies Exam	Treatment vs. Control	Yes	Yes	No	No	Yes	Yes
7	No student assessment	No student data collected	No	No	No	No	No	Yes
8	History Items Aligned with five Standards Assessed on the Nebraska Statewide Assessment and AP History Test	Treatment vs. Control	Yes	No - mean percent not mean score	No	No	No	Yes
9	No student assessment	No student data collected	No	No	No	No	No	No
10	Modified NAEP U.S. History Test, NC End of Course Test	Treatment vs. Control	Yes	Yes	No	No	No	No?
11	Modified NAEP U.S. History Test	Treatment vs. Control	Yes	Yes	No	No	No	No

Continued on the next page

Continued from the previous page

Grantee	Assessment	Comparison Type	N for each group	Mean by group	SD for each group	Effect Size	T/F Statistic	Multiple grades
12	Kentucky Statewide Core Content Test in Social Studies	Treatment vs. Control	Yes	Yes - but different tests	No	No	No	No
13	Test not specified	Treatment vs. Control	Yes	No	No	No	No	No
14	Test not specified	No student data collected	No	No	No	No	No	No
15	TAKS Statewide	Treatment vs. Control	Yes	Yes	Yes	No	No	Yes
16	California Standards Test	Treatment vs. Control	Yes	Yes	Yes	No	Yes	Yes
17	California Standards Test	Treatment vs. Control	Yes	Yes	Yes	No	No	Yes
18	NAEP	Treatment vs. Control	Yes	Yes	Yes	No	Yes	No
19	California Standards Test	Treatment vs. Control	Yes	Yes	Yes	No	No	No
20	California Standards Test	Treatment vs. Control	Yes	Yes	Yes	No	No	Yes
21	South Carolina Statewide and the AP History Exam	Treatment vs. Constructed Comparison group at the district level	Yes	No -mean percent not mean score	No	No	No	Yes
22	Long Beach DistrictWide Benchmark History Test	Treatment vs. Constructed Comparison group at the district level	No	No	No	No	No	No
23	Project created assessment	Treatment vs. Control	Yes	No	No	No	No	No
24	California Standards Test	Treatment vs. Control	Yes	Yes - Scale score	Yes	No	No	Yes

Continued on the next page

Continued from the previous page

Grantee	Assessment	Comparison Type	N for each group	Mean by group	SD for each group	Effect Size	T/F Statistic	Multiple grades
25	TCAP Statewide Achievement Test in Social Studies	Treatment vs. Control	Yes	Yes	Yes	No	Yes	Yes
26	California Standards Test	Treatment vs. Control	Yes	Yes	No	No	No	Yes
27	California Standards Test	Y1 vs. Year 2 cohort	Yes	No	No	No	No	Yes
28	Project-Based	Treatment vs. Control	Yes	No	No	No	No	Yes
29	AP History Exam	Treatment vs. Constructed Comparison group at the district level	Yes	No - mean percent, not mean score	No	No	No	Yes
30	Not specified but aligned with Nevada History Standards	Treatment vs. Control	Yes	Yes	No	Yes	Yes	No
31	TAKS Texas Statewide Social Studies Test	Treatment vs. Control	Yes	Yes	No	No	No	Yes
32	Florida Statewide	Treatment vs. Control	Yes	Yes	No	Yes	No	Yes

List of Citations for Studies

Baker, A.J. (2008). 2004 Final performance report – TAH. PR Award #U215X040316 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Black, A. (2008). 2004 Final performance report – TAH. PR Award #U215X0400897 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Brinson, J. (2008). 2004 Final performance report – TAH. PR Award #U215X040166 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Ford, M. (2008). 2004 Final performance report – TAH. PR Award #U215X040001 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Goren, G. (2008). 2004 Final performance report – TAH. PR Award #U215X040118 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Junge, J. (2008). 2004 Final performance report – TAH. PR Award #U215X040058 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Moyer, J. (2008). 2004 Final performance report – TAH. PR Award #U215X040310 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Perzan, M. (2008). 2004 Final performance report – TAH. PR Award #U215X040187 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Pesick, S. (2008).2004 Final performance report – TAH. PR Award #U215X040137 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Stewart, D. (2008). 2004 Final performance report – TAH. PR Award #U215X040339 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Wiggington, T. (2008). 2004 Final performance report – TAH. PR Award #U215X040044 Budget period #1, Report type: Final performance. Available from U.S. Department of Education, Washington, D.C. 2020-55335

Reliability of Assessments

Few of the TAH reports provided any information about the technical qualities, including the reliability, of the student assessments. Thus, it was not possible to determine which assessments had poor reliability. In the case of the statewide tests, available technical manuals were examined. The technical documentation for these assessments did not provide the actual reliability coefficients. However, because these statewide assessments are designed, developed and validated according to industry standards, it was assumed that the reliability coefficients were adequate.

Reliabilities for project-based assessments were also not reported in the TAH reports. Reliability of the NAEP American history items was not reported, as these items are not typically aggregated and reported as a single measure. Finally, the TAH report using the Newman, Bryk and Nagaoka (2001) methodology was examined to see if any information was reported about the inter-rater reliabilities associated with the scoring of student work. No such information was made available. The original article describing the methodology was examined to determine whether it provided any overall evidence of the reliability of the scoring process. Although the authors apply a systematic approach to the scoring of the student work, they do not report inter-rater reliability. The unreliability of the assignment and student work scores could be addressed in a Many Facet Rasch Analysis. This procedure is used to construct an overall measure of the intellectual quality of each assignment and adjust for any observed differences among the raters as they score comparable assignments and student work.

Combining Measures of Student Achievement

A meta-analysis requires several key judgments about the similarity of the student assessment data. Among the 12 projects included in the last stage of screening for a meta-analysis, there were four types of assessments used: statewide assessments from four different states, items from the NAEP American history test, student work samples based on the Newman, Bryk and Nagaoka (2001) methodology, and project-developed American history measures. Exhibit 6 presents the four kinds of assessments used and the number of projects that used each type of assessment.

Exhibit 6: Number and Types of Assessments Used in 12 Evaluation Reports

Assessment Type	No. of Projects
Project Developed Assessments	$N=3$
Newman, Bryk and Nagaoka (2001) Student Work Samples	$N=1$
NAEP American History Test Items	$N=1$
Statewide Assessments	$N=7$

Aggregating results across the assessment types requires that the assessments measure the same construct—in this case, student achievement in American history. The following paragraphs consider each of the four types of assessments and its relationship to learning of American history. The intent was to create a crosswalk relating the content in each type of assessment to

the NAEP American History Framework. The NAEP framework is used because in the absence of national standards in American history it offers the measure closest to a nationally recognized, objective standard. If the content in each type of assessment aligns with the dimensions of the NAEP Framework, it is reasonable to combine results from the four assessment types in the meta-analysis.

The following three Dimensions compose the core of the NAEP American History Framework:

1. **Historical knowledge and perspective**
   a. includes knowing and understanding people, events, concepts, themes, movements, contexts, and historical sources;
   b. sequencing events;
   c. recognizing multiple perspectives;
   d. seeing an era or movement through the eyes of different groups; and
   e. developing a general conceptualization of American history.

2. **Historical analysis and interpretation:**
   a. identifying historical patterns;
   b. establishing cause-and-effect relationships;
   c. finding value statements;
   d. establishing significance;
   e. applying historical knowledge;
   f. making defensible generalizations;
   g. rendering insightful accounts of the past;
   h. includes explaining issues; and
   i. weighing evidence to draw sound conclusions.

3. **Themes**
   a. change and continuity in American democracy: ideas, institutions, events, key figures, and controversies;
   b. the gathering and interactions of peoples, cultures, and ideas;
   c. economic and technological changes and their relation to society, ideas, and the environment; and
   d. the changing role of America in the world.

Project-based assessments could not be analyzed using the crosswalk with the NAEP framework. The test items and subscales comprising the project-based assessments were not available within the reports; further inquiry for the information was attempted but was unsuccessful. Thus, it was not possible to confirm their exact content and analyze them in relation to the NAEP History Framework.

Newman, Bryk and Nagaoka (2001) scoring of student work makes use of general rubrics such as "authentic intellectual work" when scoring student performance. These general rubrics were developed by the authors of the methodology and are not subject-matter specific. In this review, the student assignments and student work were focused on American history. Because teacher assignments were not provided, it was not possible to characterize the content for use in the crosswalk.

The NAEP American history assessment items were all aligned to the NAEP American History Framework and can be considered measures of American history achievement.

Four statewide assessments were used as dependent measures among the 12 projects that were included in the comparison. These statewide tests included: 1) California's California Standards Test (CST); 2) South Carolina's Palmetto Achievement Challenge Test (PACT); 3) Tennessee's Comprehensive Assessment Program (TCAP); and 4) Texas' Assessment of Knowledge and Skill (TAKS). Using the crosswalk, it was possible to determine whether test scores from the different statewide assessments could be combined as a single construct—student achievement in American history. Thus, American History Standards associated with each of the statewide assessments were related to the NAEP American History Framework.

The NAEP framework has three broad dimensions (i.e., Historical Knowledge and Perspective) followed by numerous supporting subdimensions (i.e., sequencing events). The analysis was conducted at the levels of the broad dimensions in the NAEP framework because the state standards documents revealed considerable overlap with the NAEP framework. This overlap made it unnecessary to analyze the standards content at the grain size represented in the subdimensions of the NAEP framework. For the purposes of this crosswalk, the aggregation of student test scores was done at the highest levels—in other words, at the level of the overall American history test score and not at the subdimension level. Results of the crosswalk analysis revealed that the four statewide assessments were well aligned with the NAEP Framework and could be combined for some analyses. More specific results are presented below.

Researchers conducted a crosswalk for each state relating the NAEP American History Framework to the standards by each state in American history. The dimensions of the NAEP American History Framework were identified and then compared to each state's American History Standards at the grade levels included in the TAH project reports. Below are the topical areas covered by each state's standards and the grade levels they represent:

- California :
  - Eighth-grade topics included: American Constitution and the Early Republic, The Civil War and its Aftermath;
  - Eleventh-grade topics included: Foundations of American Political and Social Thought, Industrialization and the American Role as a World Power, United States Between the World Wars, World War II and Foreign Affairs, and Post-World War II Domestic Issues.

- South Carolina :
  - Third-, fourth-, and fifth-grade topics include: History, Government and Political Science, Geography, and Economics

- Tennessee :
  - Fourth-, fifth- and eighth-grade topics include: Governance and Civics, Geography, American history Period 1, American history Period 2, and American history Period 3

- Texas :

- Eighth-, tenth-, and eleventh-grade topics include: Geographic Influences on History, Economic and Social Influences on History, Political Influences on History, and Critical Thinking Skills

Based on review of each of the state crosswalks relating the NAEP American History Framework to their standards, the only topical area represented in a state's standards that did not align with the NAEP Framework was geography in the state of Tennessee. All other topical areas represented in the state standards were related to the NAEP Framework. All of the major dimensions of the NAEP American History Framework were covered by the American History Standards associated with each state; therefore, the American history scores based on each state's assessment could be combined in a meta-analysis to represent achievement in American history.

Basis for combining across assessment types. Based on the analysis of each type of assessment, its content, and the results of the crosswalk, it was deemed reasonable to combine assessments as a single dependent variable in a potential meta-analysis.

The Department of Education's mission is to promote student achievement
and preparation for global competitiveness by fostering educational excellence
and ensuring equal access.

www.ed.gov