Evaluation of the Teaching American History Project: Bringing History Home II

Final Report, December, 2007

Submitted to:
Elise Fillpot, PhD. Project Director

Submitted by:
University of Iowa Center for Evaluation and Assessment
Julie Kearney, Emily Lai, and Don Yarbrough, Director
210 Lindquist Center
Iowa City, IA 52242
Email: coe-cea@uiowa.edu
Website: http://www.education.uiowa.edu/cea/

The Bringing History Home II Project is supported by a grant award to Washington Community Schools, Elise Fillpot, project director, from the United States Department of Education Teaching American History Program. This evaluation was conducted under the auspices of a sub-contract between Washington Community Schools and The University of Iowa Center for Evaluation and Assessment and supported by funds from the above grant, number U215X030089.

Table of Contents

Table of Tables

Table of Figures

The Year 3 *Bringing History Home II Project* evaluation focused on five areas: 1) monitoring the grant processes and products during Year 3, 2) providing formative feedback on ways to continue to improve the processes and products, 3) collecting survey and focus group data from teachers on their second and third implementations of the *BHH* curriculum, 4) observing classrooms during *BHH* curriculum implementations, and 5) collecting multiple forms of written assessments of student outcomes and conducting focus groups with participating students in 3rd through 5th grades.

The University of Iowa Center for Evaluation and Assessment (CEA) is the third party evaluator for this project. Its director and staff drafted all evaluation designs and instrumentation, shared the drafts with the *BHHII* Project Director and selected project staff and stakeholders for suggested revisions, and then revised as needed. At all times, significant groups of stakeholders, including the project partners and selected teachers, had opportunity to provide critiques of the evaluation materials and approaches.

Throughout the three project years, at least one member of the evaluation team attended and observed all project team meetings and steering committee meetings and provided on-going informal evaluative comments about ways to improve project processes and products, including team member collaboration and interactions. Other team members, including several PhD graduate students, also assisted with various aspects of instrument development, evaluation design, information collection, focus group facilitation, group interviews, and analysis.

1. Executive Summary

This report is the final evaluation report for the *Bringing History Home II Project (BHHII)* that received funding in 2003-06 as part of the U.S. Department of Education *Teaching American History* program. The Center for Evaluation and Assessment (CEA) at the University of Iowa evaluated this project (and the previous *Bringing History Home Project (BHH))* throughout the grant period as a third-party evaluator. Evaluation methodology included teacher surveys, teacher focus groups, classroom observation, student written assessments, and student group interviews. Data collected through these methods were analyzed and summarized and summaries are included in the Appendices to this report.

The purposes of the BHH II project evaluation were to:

1) Contribute to project improvement
2) Document the actual activities and procedures of the project as implemented
3) Investigate changes over the three years in teachers' attitudes concerning teaching history and their implementation of history curricula
4) Examine students' historical content knowledge and historical thinking skills.

During the BHH II grant period, K-5 students in four Iowa school districts received history instruction using the BHH curricular units. Approximately 123 teachers taught history to elementary students using the BHH curricula and nearly 3000 students learned history each year of the project. Students in three districts received BHH history instruction for two consecutive years and students from the original BHH district continued to receive instruction in the BHH curricula so that by the conclusion of this project, about 100 students had taken part in the curricula for as many as four and a half years. The BHH II project provided professional development workshops for participating teachers during the first two summers of the project and ongoing professional development support during the entire grant period. The BHH II professional development model was in keeping with research showing that teacher professional development is most effective when it provides teachers with new content knowledge, promotes teaching new curricula as a cohesive part of other teaching activities, occurs at the school (or district) level, and is of sufficient duration (Garet, et al., 2001). The project's professional development activities that took place during Years 1 and 2 have been described in detail in previous annual reports submitted to the U.S. Department of Education and these reports are available at the CEA website at: http://www.education.uiowa.edu/cea/tah/index.html.

During the final year of the project, the primary focus of the BHH II project evaluation was to document and describe teacher and student outcomes. Teacher survey data showed that

teachers who participated in the BHH II project were more likely to teach U.S. history in their classrooms (particularly, but not exclusively, historical content areas addressed by the BHH curriculum) and rated their students' knowledge of history and ability to think historically higher than did their comparison group peers. Treatment group teachers were also likely to believe that it was beneficial for their students to learn history in the elementary classroom setting, and were more likely than their comparison peers to believe that using pedagogical practices associated with fostering historical thinking skills (e.g. discovery learning, guided inquiry, primary source analysis) is useful in elementary classrooms.

When asked about their self-efficacy in teaching history, treatment and comparison group teachers appeared to be nearly equally comfortable teaching history, in their enjoyment of teaching history, and in their confidence in using primary sources to teach history. However, it is difficult to interpret these findings because elsewhere, comparison teachers said they taught very little history and appeared unsure about what defined a primary source. In addition, when asked to list the history topics they taught, comparison teachers often named topics that are not actually *history*, such as state capitals, regions, voting, or national holidays. Comparison teachers also frequently listed a large number of history topics taught, but said that they spent less time teaching these topics, indicating a surface treatment of the content. When asked to describe primary sources they had used in their teaching, comparison teachers (particularly those in the younger grades) often listed things that are not primary sources, such as textbooks, globes, worksheets, and picture books. On the other hand, through participation in the BHH II project, treatment teachers appear to have developed an informed philosophy of what it means to teach history and of how they can facilitate students' acquisition of the skills required to think historically.

The second primary emphasis of the BHH II project evaluation during the final year was to collect data showing student outcomes. Assessments were piloted in treatment schools during previous years of the BHH and BHH II projects, so during the final year, written assessments were collected from all 3rd through 5th grade treatment and comparison students, and group interviews were conducted with small samples of 3rd through 5th grade treatment and comparison students. All 3rd through 5th grade students completed one type of written assessment aligned with a particular historical era (eras that were addressed by the BHH curricular units). The assessments were either: 1) narratives students constructed using seven key words from the unit, or, 2) photograph analyses using a photograph from a historical era students in treatment classrooms had studied as part of the BHH curricula. Some students (randomly assigned at the classroom level) completed both pretests and posttests, and some students completed the posttest

only. There were therefore twelve different assessments administered to BHH II students: narrative assessments relevant for each of three grade levels and for each of the two grade-level units, and photo analysis assessments, also specifically designed for each grade level and unit.

Two major trends were observed on all twelve assessments. First, the mean performance of students in treatment classrooms improved from pretest to posttest condition, indicating that students had learned the necessary historical content and skills. On all six narrative assessments treatment posttest means were significantly greater than pretest means (as evidenced by non-overlapping 99% confidence intervals). On the photo analyses assessments, treatment posttest means were all greater than treatment pretest means and in five of the six unit assessments, mean score improvements also differed significantly using 99% confidence intervals. Conversely, comparison students' performance did not significantly improve from pretest to posttest on any of the twelve narrative or photo analyses assessments.

The second trend was that treatment students outperformed comparison students. Treatment students' mean performance on nearly all assessments was significantly greater (as evidenced by non-overlapping 99% confidence intervals) than their comparison group peers. This means that students in treatment classrooms were learning historical content knowledge and skills that were not learned by students in comparison classrooms. Treatment group posttest mean scores were significantly higher than comparison students' posttest mean scores for all narrative assessments. On the photo analyses assessments, treatment students' posttest mean scores were higher than their comparison peers' posttest mean scores in all cases, and in five of the six cases treatment and comparison means were significantly different using 99% confidence intervals.

In addition, during student group interviews conducted with small heterogeneous samples of students, treatment students revealed deep oral knowledge of the historical content studied, the ability to apply background knowledge and learn new information from historical photographs to construct plausible narratives about events in history, and genuine excitement about learning history in general. In contrast, during comparison group interviews, comparison students displayed very little historical content knowledge and did not demonstrate skills necessary to learn new information from novel historical photographs.

The BHH II project took the success of the pilot project BHH and established that the BHH curricula could be used to great effect in teaching historical content and historical thinking skills to students in K-5 classrooms in four Iowa districts. The curricula have been adopted in all project schools and either the curricula and/or the pedagogical methods ascribed by the BHH projects have been or will soon be adopted by additional districts in four other states. The result

for teachers and students has been evidence to foster the successful expansion of the role of U.S. history in elementary classrooms.

2. Project Description

The *Bringing History Home II Project (BHHII)* is a three-year project of the Teaching American History Program funded by the U.S. Department of Education. The project is an extension of an earlier project, *Bringing History Home (BHH)*, which was completed in 2004. The original project took place in the two primary grade schools (one a K-3 school and the other a 4-6 school) in Washington, IA. The *BHHII* grant expanded the project to include three new districts in Iowa; Creston, Perry, and Maquoketa. In addition to the three new districts, the Washington Community School District (WCSD) adopted the curriculum as part of their district's social studies curriculum and continued to teach the *BHH* curriculum during the *BHH II* grant period. The *BHH II* project provided participating schools with 1) Professional development workshops on teaching history to elementary school teachers, 2) Written history curricular units (at least two for each grade), 3) Resources to support the teaching of the units (including books, videos, other supplies), and 4) Ongoing professional support from the project director, project staff, and teacher mentors during the implementation phase.

Although staff numbers at each school varied naturally over the course of several years, there was no teacher attrition due to factors other than typical mobility, retirement, etc. The typical number of teachers participating over the three project years was 99 teachers in the three *BHHII* schools and an additional 24 teachers continuing to participate from the Washington schools. Student attendance over the same time period also naturally fluctuated, but there was no unexpected attrition or additions to the student bodies. Table 1 provides the 2005-06 (Year 3) totals and demographic characteristics of students in *BHHII* participating schools (including Washington) and in the two schools used for comparison purposes for the project evaluation. The number of students in the three *BHHII* schools for Year 3 was 2047. An additional 876 students continued to learn the *BHH* curriculum in the WCSD for a total of 2923 students experiencing the curriculum during 2005-06.

The project activities that took place during the first two years of the grant were described in detail in the Year 1 and Year 2 reports which were submitted to the US Department of Education and they continue to be available on the CEA website at: http://www.education.uiowa.edu/cea/tah/index.html.

During the third year, grade level groups at all schools were invited to receive on-site support for their curriculum implementations through visits from project staff. During the 2005-06 school year, 16 of the 18 possible grade level groups worked during the school day or after school with project staff to prepare for and become more comfortable with teaching their *BHH*

units. A few teachers also received on-site one-on-one help from project staff. Email and telephone support from project staff and mentor teachers was also available. There were no additional large group workshops or meetings conducted during the third year of the grant period.

Table 1 Demographic Characteristics of BHHII Treatment and Comparison Schools, 2005-06.

K-5 Schools		Racial/Ethnic Identity						Number of English Language Learners (LEP)	Percent Eligibility for FRP (School %)
		White	Black	Asian	Hispanic	Native American	K-5* TOTAL	K-5	K-5*
Treatment Schools	District A	555	19	15	12	2	603*	0	55%*
	District B	633	10	9	12	3	667*	9	52%*
	District C	419	6	8	384	0	817	212	64%
	District D	741	26	6	111	1	885*	32	38%*
Comparison Schools	District E	309	11	4	7	0	331	7	40%
	District F	886	4	1	4	3	898* **	0	28%* **

* These schools include pre-K and statistics provided for ethnicity and eligibility for Free or Reduced Price Lunch include pre-K students. The numbers of pre-K students in these districts are as follows District A=25, District B=15, District D=9, District F=17.
** This school includes 6th grade and statistics provided for ethnicity and eligibility for Free or Reduced Price Lunch include 6th grade students. The number of 6th grade students in this school = 120.

Source: *Iowa Department of Education* Education Statistics Web Page:
http://www.iowa.gov/educate/statistics/education/data/student-data/
Files on Enrollment, English Language Learners, and Free and Reduced Price Lunch 2005-06

During the third year, all teachers completed their second and third curriculum implementations, teaching one unit for the first time and another unit for the second time. Nearly all the teachers chose to teach one unit during the fall of 2005 and the other in the spring of 2006, but teachers were allowed to teach the curriculum whenever they chose, and there were a few variations in the timing and sequence in which the curriculum units were taught, described more fully in Table 2.

The *BHH* curriculum units and when they were taught by most teachers can be found in Table 2.

Table 2. Schedule of Implementation of BHH Units 2005-06*

	Spring 2005	Fall 2005	Spring 2006
Kindergarten	History of Me	History of Me	Children Long Ago
Grade 1	My 1st Grade History	My 1st Grade History	Community History
Grade 2	Immigration	Immigration	Environmental History
Grade 3	Industrialization	Industrialization	Segregation
Grade 4	Great Depression	Progressive Era	Great Depression
Grade 5	WWII Home Front	Native American History	WWII Home Front

*Exceptions to the schedule shown in Table 2: One school taught the 2nd grade Native American unit (not evaluated) instead of the Immigration unit because their school already had an immigration unit in place at the third grade. One school taught the 1st grade Community History unit in the fall of 2005 and the 1st Grade History Unit in the spring of 2006 and another school taught the 4th grade Great Depression unit in the fall of 2005 and the Progressive Era unit in the spring of 2006.

3. Evaluation Methodology

For the final year of the evaluation of the *Bringing History Home II Project*, five data collection methods were used: 1) Teacher Implementation Surveys, 2) Student Written Assessments, 3) Classroom Observations, 4) Teacher Focus Groups, and 5) Student Group Interviews. Evaluation team members also attended all organizational meetings for the grant process and communicated regularly with the project director concerning the curriculum implementations, workshop development, plans for dissemination, website updates, and grant partner work. [Data collection methods used in previous grant years have been described in detail in previous reports available at the CEA website given above.]

3.1 Teacher Implementation Surveys

During the spring of 2005, fall of 2005 and spring of 2006 evaluators asked participating teachers from the three *BHHII* school districts, the original *BHH* school district, and two comparison school districts to complete surveys concerning their teaching of history, either through implementation of the *BHH* curriculum or their regular history curriculum. Surveys were mailed to the teachers or given to them in person by the evaluator. Teachers were given postpaid envelopes to return the survey. During the 2005-06 school year, the surveys were given to all teachers prior to their curriculum implementations so that teachers could complete the items concerning implementation as they completed each activity in the unit. The surveys were all designed to be used in conjunction with a particular *BHH* unit and consisted of a combination of scaled and open-ended items concerning 1) the teachers' self-described thoroughness of teaching different elements of the *BHH* curriculum, 2) teachers' perception of their students' competence at performing different skills or student knowledge of content related to the curriculum, 3) teachers' opinions of the benefits of the curriculum for their students, 4) teachers' use of different pedagogic techniques during their implementation of the curriculum, 5) teachers' use of primary sources in teaching history, and 6) teachers' opinions on teaching history and their opinions about their own and their perception of students' general interest and comfort with learning history.

Twelve unique surveys were designed to align with the content and the goals of each of the 12 *BHH* curricular units. In addition, the survey administered during the Fall of 2005 was shortened by omitting some of the open-ended items. The omitted items were part of the survey in both Spring 2005 and Spring 2006 so teachers' opinions at two different points in time could be ascertained. A copy of each of the surveys is the in the Appendices to this report. Because *BHHII* participating teachers taught the same topic twice during the grant period, for each grade

level, two surveys covered the same content, as taught on successive implementations, and the other was concerned with the topic that teachers each taught only once (See Table 2).

Because of ordinary changes in the staffing numbers in the schools, the number of teachers in treatment schools surveyed varied from 117 to 121 over the three cycles of data collection. For the Spring 2005 survey, the response rate was 84%, for the Fall 2005 survey the response rate was 97%, and for the Spring 2006 survey the response rate was 95%. In the comparison group, 40 of the 45 teachers who were sent surveys replied for a response rate of 89%. The comparison teachers who did not return the survey during Spring 2005 were not surveyed during the next school year and one comparison teacher discontinued participation in the project because of a low response rate from students on human subjects consent forms, so the response rate during Fall 2005 and Spring 2006 was 98% of teachers surveyed.

3.2 Student Written Assessments

During Year 3, written assessments in three different forms were collected from all student participants in 3rd through 5th grades. Evaluators developed the instruments with the collaboration of Washington mentor teachers and the project director. Students from the WCSD piloted some of the instruments during the first two years of the grant period. For each unit, students completed only one type of assessment, so for any particular unit, data was collected with each assessment type from approximately one-third of the students. Sampling was done at the classroom level, so students within a particular classroom all completed the same type of assessment. In addition, assessment conditions varied with a randomly selected half of the classrooms completing pre and post-assessments (of the same type) for each unit and the other half completing post-only assessments so that practice effects could be examined. Classrooms were randomly assigned using stratified random assignment (by school) to assessment types and conditions for the first unit and then assigned to a different type and administration condition for the second unit. Data was collected twice (once for each unit) during the 2005-06 school year but conditions were manipulated so that students did not repeat assessment type (for the second unit) and were only pre-tested once. Table 3 shows the distribution of the assessment types and condition of assessment across the participating classrooms. The distribution is not entirely evenly distributed because of odd numbers of classrooms and because of changes in numbers of classrooms after the plan was established and materials prepared.

Teachers were responsible for administering all student assessments to their classrooms. They were provided with assessments for all students, detailed instructions for administration including a brief script to be used for test administration, and postage-paid envelopes for

returning completed materials. Teachers administered pre-assessments before beginning any instruction in the topic area and administered the post-assessments as soon as possible after they had completed their instructional units. They were asked to indicate the date of the test administration, the number of students taking the assessments, and the number of the students in the classrooms. They were also invited to comment on the assessment.

Table 3. Number of BHH II Classrooms Assigned to Each Assessment Condition

Grade/ Condition	Unit 1						Unit 2					
	Instrument Type						Instrument Type					
	Narrative		Photo Analysis		What is History?		Narrative		Photo Analysis		What is History?	
	Pre/ Post	Post only	Pre/ Post	Post only	Pre/ Post	Post only	Pre/ Post	Post only	Pre/ Post	Post only	Pre/ Post	Post only
3/Treatment n=18/19*	3	3	4	3	3	2	3	4	3	4	1	4
3/Comparison n=6	2	1	1	1	0	1	1	1	1	2	1	0
4/Treatment n=20**	4	3	4	4	2	3	3	4	4	3	2	4
4/Comparison n=8	1	1	1	2	1	2	2	1	2	1	1	1
5/Treatment n=20**	3	4	3	3	3	4	4	4	5	3	1	3
5/Comparison n=4	1	1	1	1	0	0	1	1	0	1	1	0

*One classroom did not do first 3rd grade unit.
**In one school, one teacher in each of 4th and 5th grades teaches 5 sections, so they were given materials to do five different conditions.
Classrooms were randomly assigned to conditions for Unit 1 and then to a different assessment type and condition for Unit 2.

The three types of assessments used for data collection during Year 3 were:

3.2.1 Narrative Assessments

Students in classrooms that were assigned to the narrative assessment condition were asked to write a narrative incorporating six or seven key terms from the history unit in which they had received instruction. Students were asked to use the terms to tell a story about something that happened in the past. Teachers were familiar with this format of assessment because it was introduced as a potential means of assessment during summer teacher development workshops. Some teachers reported using similar instruments (although not necessarily employing the same terms) during the regular course of their instruction.

Scoring of Narrative Assessments

The rubric used to score student responses was constructed in several stages. Evaluators first constructed "acceptable" definitions and contexts for usage from the *BHH* written

curriculum, including topic specific glossaries provided by the project. Evaluators selected examples of student responses illustrating each point of the score scale for each term. The scale for rating the correctness of students' usage of each term ranged from 0 to 2, with "0" indicating no answer or an incorrect answer, a score of "1" indicating a minimally or partially correct answer with little or no elaboration, and a score of "2" for a correct answer with some elaboration.

Two evaluators independently scored small samples of student responses and then met to compare their scores. Where there were disagreements, a consensus approach was utilized. When the inter-rater agreement between the two evaluators (measured by percent exact agreement) reached levels of at least 90%, the rubric was considered complete, and training materials were finalized. A trained rater scored all student responses. During training, the rater met with evaluators and went over several small samples of student responses to further illustrate how to apply the rubric. The rater then independently scored small samples of student responses. Evaluators monitored the extent of agreement and made corrections and modifications in scoring procedures where necessary.

The inter-rater agreement between the trained rater and the evaluator consensus scores was calculated for a sample of narratives throughout each set and is found in Table 4. In calculating percent agreement, the scores were counted as *different* if they were not exactly the same, so that a disagreement between a score of "1" by one coder and a score of "2" by the other , was treated the same as if they had been scored "0" and "2". The total number of scores for which the two coders had complete agreement was divided by the number of possible scores to calculate the percent agreement. If percent *adjacent* agreement were used instead of *exact* agreement, this figure would have been larger since most differences were between scores of "1" and "2". Complete instructions for rater training for each of the six units evaluated are included in the appendices of this report.

Table 4. Inter-rater Agreement on Coding of Student Narrative Assessments

Grade and topic	Number of training and reference cases		% Agreement
	Narratives	Terms	
3rd Grade			
Segregation	38	7	88%
Industrialization	51	6	89%
4th Grade			
Great Depression	31	7	87%
Progressive Era	66	7	94%
5th Grade			
Native American History	23	7	84%
WWII Home Front	52	7	93%

Reliability of Aggregated Scores

In order to summarize the internal consistency of the narrative assessments, we computed coefficient alphas for all items (seven items on most assessments, on one assessment, six items) on pre and posttest conditions, summing together the treatment and comparison scores on the same assessments. Table 5 presents the internal consistency reliability estimates (coefficient alphas) for the aggregated scales at pre and post testing. Overall alpha values for the posttest condition range between 0.72 and 0.87.

These estimates reflect several sources of error that tend to reduce the reliability of performance assessments of this type. First, there are a limited number of separately scoreable elements on the assessment. In this case, students had the opportunity to earn points on only seven terms (six on one assessment). Second, the scoring rubric applied to the assessment resulted in a maximum of only two points possible for each term. This means that total scores (summed across all terms) could only range from 0-14 (0-12 for one assessment), producing minimal score variation, which tends to attenuate reliability estimates. Scoring rubrics involved trained raters, which introduced subjectivity (and some amount of unreliability) into the scores. Finally, characteristics of the experimental conditions tended to further attenuate estimates of the reliability of the instrument, such as a (predictable) lack of variability in scores for comparison students (on both pre- and post-tests) and treatment students (primarily on the pre-tests). (Linn & Miller, 2005, Mehrens & Lehmann, 1991). For this type of measure, alpha coefficients of 0.80 and above are generally seen as acceptable. Educators who prefer constructed response assessments over those relying primarily on multiple-choice items believe the slight decrease in reliability is at least partially offset by an increase in construct validity.

Table 5. Internal Reliabilities for Student Narrative Assessments -- Coefficient Alphas

Grade/Unit Test Condition (n)*		Alpha	Standardized Alpha
3rd Grade Industrialization (k=6)**			
Pretest	(69)	0.63	0.62
Posttest	(102)	0.79	0.79
3rd Grade Segregation (k=7)			
Pretest	(41)	0.32	0.50
Posttest	(113)	0.76	0.76
4th Grade Great Depression (k=7)			
Pretest	(42)	0.38	0.31
Posttest	(95)	0.83	0.83
4th Grade Progressive Era (k=7)			
Pretest	(87)	0.66	0.58
Posttest	(130)	0.72	0.72
5th Grade Native American History (k=7)			
Pretest	(64)	0.56	n/a
Posttest	(125)	0.75	0.75
5th Grade WWII Home Front (k=7)			
Pretest	(35)	0.86	0.86
Posttest	(121)	0.87	0.86

*n = number of students completing assessments
**k = number of items on assessment

Item Discriminations and Difficulties of Narrative Assessments

For each grade and unit, test item data were analyzed to determine item discrimination for each item, expressed as the correlation between each item and the total score on the assessment. In addition, test item data were also analyzed to determine the difficulty of each item expressed as the mean score of the item for each group taking the assessment, with a maximum of two points possible on each item. The full tables of item data for each testing condition (pre and post, treatment and comparison) are included in the appendices to this report. Table 6 shows selected details concerning item discrimination and difficulty on narrative assessments.

Posttests item discriminations are shown for treatment students since that is the group for whom the test was most likely to have the largest variance in scores and the concept of discrimination is most relevant. As can be seen, for most items, discrimination was quite good for open-ended items, with 36 of the 41 items across all tests having item discrimination of 0.30 or larger, and 27 items having discrimination indices greater than 0.40. Two items had particularly poor discrimination – both on the 3rd grade Industrialization test – with item discriminations of -0.01 for *factory conditions* and 0.17 for *age of inventions*. Since these items were both part of the same test, it is possible that third grade scores are in general less stable due

to typically lower writing skills. This test also had only six items instead of seven, which may have led to lower item discrimination indices.

In regards to item difficulty, all items except one were more difficult for comparison students than they were for treatment students, and all items were less difficult for treatment students at the posttest than at the pretest. The only item on which the comparison students scored slightly higher than treatment students was the word *slavery*. The concept of slavery is a common topic for instruction in multiple contexts in grade school during Black History Month. Comparison students, for whom the other items on the 3rd Grade Segregation assessment were quite difficult, may have spent more time on that item and earned closer to the full value of points available, while treatment students' attention was spread more evenly across the other items. Items on other tests that were about equally easy for comparison and treatment students included *immigration* and *dust bowl*, both of which are concepts commonly studied in a social studies curriculum, while the other terms present in the same tests as these were less familiar and more difficult for comparison students.

Table 6. Item Discrimination and Difficulty for Narrative Assessments

Grade/Topic Item	Item Discrimination	Item Difficulty *(Mean score with a range of 0-2 possible)*		
	Treatment posttest	Treatment Pretest	Treatment Posttest	Comparison Posttest
3rd/Industrialization				
Craftsman	0.32	0.44	1.37	0.46
Single owner	0.40	0.36	1.31	0.24
Partnerships	0.34	0.47	1.32	0.30
Age of inventions	0.17	0.24	1.37	0.32
Corporations	0.62	0.11	0.91	0.19
Factory conditions	-0.01	0.04	1.09	0.78
3rd/Segregation				
Slavery	0.47	0.77	0.98	1.17
Constitution	0.48	0	0.56	0.52
Amendment	0.62	0	0.66	0
Thirteenth Amendment	0.68	0.03	1.10	0.04
Segregation	0.37	0.13	0.88	0.43
Prejudice	0.59	0.07	0.52	0.22
Jim Crow Laws	0.29	0.19	0.58	0.09
4th/Great Depression				
Roaring 20's	0.29	0.04	0.83	0.03
Great Depression	0.47	0.33	1.32	0.23
Dust Bowl	0.32	0.25	1.09	0.73
Penny auctions	0.40	0.08	0.74	0.37
Migrant worker	0.50	0	0.68	0
Hooverville	0.41	0	1.18	0
Unemployment	0.40	0.21	0.91	0.10
4th/Progressive Era				
Child labor	0.20	0.38	1.51	0.71
Tenement housing	0.32	0.35	1.26	0.78
Capitalism	0.48	0.30	0.61	0.29
Profit	0.58	0.35	0.76	0.38
Immigrant	0.42	0.58	1.12	1.08
Progressives	0.32	0.10	0.35	0.04
Muckrakers	0.45	0.02	0.96	0.25
5th/Native American History				
Policy	0.30	0.14	1.01	0.35
Trail of Tears	0.52	0.17	1.33	0.75
Reservation	0.41	0.24	1.04	0.78
Assimilation	0.29	0	0.64	0
Boarding schools	0.54	0.17	1.32	0.08
Indian Reorganization Act	0.25	0	0.20	0
Pow-wow	0.43	0.17	1.11	0.53
5th/WWII Home front				
Home front	0.50	0.33	0.91	0.23
Ration books	0.49	0.13	0.96	0.03
Victory garden	0.54	0	1.08	0.03
War bonds	0.47	0.07	0.56	0.03
Internment camps	0.49	0.27	1.34	0.03
War effort	0.43	0.13	0.63	0.10
Rosie the Riveter	0.53	0.20	0.98	0.10

3.2.2 Photo Analysis Assessments

Students were asked to answer several questions in writing in reference to a photograph that was relevant to the content of the unit that they had been taught as part of the *BHH* curriculum. The *BHH* curriculum emphasizes the use of primary sources in learning history and in particular on using photographs to learn about the era of study. During professional development workshops and in the *BHH* written curriculum, age appropriate adaptations of the National Archives and Records Administration (NARA) format for doing photo analysis are suggested for use at all grade levels. Students (3rd through 5th grades in particular) were expected to be comfortable with the process of closely examining photographs, and to be able to use skills related to photo analysis combined with using relevant background knowledge to make informed speculations about a photograph they had not seen before. In 2003-2005, assessments were piloted with students in the WCSD to determine the types of questions that students were most able to answer using unfamiliar photographs concerning familiar content. The questions used for the assessments, across all three grades and all six content areas assessed, were:

- What do you think is happening in the photo?
- Who do you think the people in the photo are?
- When do you think the photo was taken?
- Where do you think the photo was taken?
- Why do you think the photo was taken?
- After you look at the photo, what questions do you still have about the photo that you would like to learn the answers to?

Students were told that if they were not sure about the answers, it was okay to guess.

Scoring of Photo Analysis Assessments

The rubric used to score student responses was constructed in several stages. Evaluators in consultation with experts first constructed acceptable responses to the "what", "who", "when", and "where" questions based on the information available concerning each photo. When scoring the photo analyses, raters used each student's responses to all questions to decide on the score because in piloting the instrument, we observed that students often gave information about the photo in a place other than for the question where it was specifically requested. (For example, a student might respond to the "What" question for a civil rights era photo, that it was probably taken during segregation times and not repeat that answer when asked "when".) Evaluators selected examples of student responses and showed how each example should be scored, covering as many points of the score scale as possible. Since it was not expected that students would actually know the exact details for the photographs (the photographs did not depict famous

people, places, or events) the answers were to be judged on their plausibility within a particular historical context. The scale for rating the plausibility and accuracy of students' usage of each term ranged from 0 to 2, with "0" indicating no answer or an incorrect answer, a score of "1" indicating a minimally or partially correct answer with little or no elaboration, and a score of "2" for a correct answer with clear and correct elaboration. Plausibility was based on expert judgments of what the photos depicted from the historians and project staff.

In addition to scoring the students' responses to the "what", "who", "when", and "where" questions, evaluators coded two additional factors for each response. First, codes were assigned indicating whether, and in which historical context the student placed their responses. This procedure was intended to establish whether the student used the cues to place the photograph into the context that they had studied and to examine the types of incorrect but somewhat plausible responses that students may have provided. (For example, some students placed a photo of Japanese Americans waiting to go to internment camps into a WWII context, but failed to note the internment camp context.) Evaluators also coded the students' responses to the "why" question into categories such as, for historical purposes, to pursue activist goals, or for more typical reasons, e.g. they were family members or a famous person.

Two evaluators independently scored small samples of student responses and then met to compare their codes and scores. Where there were disagreements, a consensus approach was utilized. When the inter-rater agreement between the two evaluators (measured by percent exact agreement) reached levels of approximately 90%, the rubric was considered complete, and training materials were finalized. A trained rater scored all student responses. During training, the rater met with evaluators and went over several small samples of student responses to further illustrate how to apply the rubric. The rater then independently scored small samples of student responses. Evaluators monitored the extent of agreement and made corrections and modifications in scoring procedures where necessary.

The percent agreement between the trained rater and the evaluator consensus scores was calculated for a sample of photo analyses throughout each set and is found in Table 7. In calculating percent agreement, the scores were counted as different if they were not exactly the same, so that a disagreement between a score of "1" by one coder and a score of "2" by the other , is treated the same as if they had been scored "0" and "2". The total number of scores for which the two coders had complete agreement was divided by the number of possible scores to calculate the inter-rater agreement. If percent *adjacent* agreement were used instead of *exact* agreement, this figure would have been larger since most differences were between scores of "1" and "2".

The coding of the historical context was also included in the calculation of the inter-rater agreement. Because placing the photo in an incorrect historical context affected the scoring of the other items, the item scores were to some extent dependent, as is often characteristic of performance assessments, which typically reduces reliability.

Table 7. Inter-rater Agreement on Coding of Student Photo Analyses Assessments

Grade and topic	Number of training and reference cases		% Agreement
	Narratives	Terms	
3rd Grade			
Segregation	46	5	92%
Industrialization	49	5	82%
4th Grade			
Great Depression	57	5	86%
Progressive Era	46	5	90%
5th Grade			
Native American History	45	5	87%
WWII Home Front	43	5	91%

Complete instructions for rater training and the photo analyses assessment forms for each of the six units evaluated are included in the appendices of this report.

Reliability of Aggregated Scores

In order to summarize the internal consistency of the scores on each photo analysis assessment, we computed estimates of reliability in the form of coefficient alphas for four items (the historical context codes were not included) on pre and posttest conditions, summing together the treatment and comparison scores on the same assessments. Table 8 presents the internal consistency reliability estimates (coefficient alphas) for the aggregated scales at pre and post testing. Overall alpha values range from 0.77 to 0.97. As with the narrative assessments, although alpha levels of 0.80 and above can be seen as acceptable for this type of measure, several factors that influence and attenuate estimates of reliability are inherent in this measure, including, test length, group homogeneity (which causes low variability), and scoring subjectivity (Linn, et al., 2005; Mehrens, et al., 1991).

Table 8. Internal Reliabilities for Student Photo Analysis Assessments Coefficient Alphas

Grade/Topic Test condition (n)*		Alpha (k=4)**	Standardized Alpha
3rd Grade/Industrialization			
Pretest	(66)	0.91	0.93
Posttest	(127)	0.77	0.78
3rd Grade/Segregation			
Pretest	(38)	0.97	0.99
Posttest	(95)	0.93	0.93
4th Grade/Great Depression			
Pretest	(89)	0.95	0.96
Posttest	(108)	0.93	0.93
4th Grade/Progressive Era			
Pretest	(48)	0.96	0.97
Posttest	(118)	0.87	0.88
5th Grade/Native American History			
Pretest	(40)	0.88	0.88
Posttest	(86)	0.95	0.95
5th Grade/WWII Home Front			
Pretest	(54)	0.93	0.94
Posttest	(113)	0.91	0.91

*n = number of students completing assessments
**k = number of items on assessment

Item Discriminations and Difficulties of Photo Analysis Assessments

For each grade and unit, test item data were analyzed to determine item discrimination for each item, expressed as the correlation between each item and the total score on the assessment. In addition, test item data were also analyzed to determine the difficulty of each item expressed as the mean score of the item for each group taking the assessment, with a maximum of two points possible on each item. The full tables of item data for each testing condition (pre and post, treatment and comparison) are included in the appendices to this report. Table 9 shows selected details concerning item discriminations and difficulties on photo analysis assessments.

Posttests item discriminations are shown for treatment students since that is the group for whom the test was the most likely to have the largest variance in scores and where discrimination is most relevant. As can be seen, for most items, discrimination was quite good for open-ended items, with 22 of the 24 items across all tests having item discrimination of 0.50 or larger and 15 items having discrimination indices greater than 0.70. Two items had somewhat poor discrimination – both were items asking the students to designate *when* the photographs were taken – with item discriminations of 0.17 for the 3rd grade Industrialization photo and 0.25 for the 4th grade Progressive Era photo.

For item difficulty, all items except one were more difficult for comparison students than they were for treatment students, and all items were easier for treatment students at the posttest than at the pretest. The only item on which the comparison students scored slightly higher than treatment students was for the item mentioned above that asked 3rd graders when they thought a photo of children laboring in a mill might have been taken and it was a difficult question for both comparison and treatment groups (mean=0.52 and 0.42, respectively). The only other item on which comparison group posttest performance approached treatment group performance was a relatively easy item, the *who* question on the same photograph, for which a correct answer only needed to designate that the people depicted were children, which was quite apparent in the photo.

The 5th grade Native American assessment in general appeared to be the most difficult for students with item difficulties on the treatment posttest ranging only from 0.10-0.64. In most cases, teachers said that they had not been as thorough when teaching this unit and said that it was difficult for their students, although treatment students showed improvement from pretest to posttest and scored higher than their comparison group peers.

In most cases, particularly in 4th and 5th grades, treatment group students also scored higher on the photo analysis pretest than the comparison group students. (Any possible differences in general ability between the two groups are described in the Results section of this report; in all cases comparison group students were generally at least or more able on measures of general ability than their treatment group peers.) The photo analysis task was designed to test not only students' content knowledge but also their ability to learn from and apply background knowledge to historical sources, in this case, photographs. Therefore, one explanation for treatment students' superior performance on the pretest might be that they already had some of the necessary analysis skills from previous years' instruction and practice in analyzing photographs and were more able to generate sensible conclusions about the photograph, even prior to formal content instruction.

In the case of the Progressive Era unit photograph (also a photo of children working in a mill, but not the same photograph used in the 3rd grade assessment), prior knowledge from a previous years' BHH unit might also have been available to help them analyze the photographs. The 3rd grade Industrialization unit touches on the topic of child labor and then it is revisited during the 4th grade study of the social problems that prompted some of the reforms of the Progressive Era. Since nonetheless there were large gains from pretest to posttest on that test as well, it is clear that their performance continued to improve with additional instruction.

Table 9. Item Discrimination and Difficulty for Photo Analysis Assessments

Grade/Topic Item	Item Discrimination	Item Difficulty *(Mean score with a range of 0-2 possible)*		
	Treatment posttest	Treatment Pretest	Treatment Posttest	Comparison Posttest
3rd/Industrialization				
What	0.76	0.36	1.06	0.47
Who	0.68	0.64	1.53	1.04
When	0.13	0.42	0.42	0.52
Where	0.71	0.36	1.04	0.52
3rd/Segregation				
What	0.80	0.07	1.13	0.38
Who	0.82	0.15	1.12	0.38
When	0.72	0.15	0.74	0.30
Where	0.79	0.11	1.00	0.22
4th/Great Depression				
What	0.76	0.16	1.03	0.29
Who	0.81	0.16	1.06	0.20
When	0.73	0.19	1.10	0.27
Where	0.79	0.10	1.11	0.22
4th/Progressive Era				
What	0.67	0.54	1.55	0.38
Who	0.56	0.65	1.74	0.50
When	0.25	0.65	0.98	0.45
Where	0.67	0.57	1.55	0.30
5th/Native American History				
What	0.95	0.23	0.62	0.10
Who	0.95	0.30	0.64	0.08
When	0.69	0.30	0.10	0.05
Where	0.91	0.17	0.60	0.05
5th/WWII Home front				
What	0.82	0.37	1.26	0.17
Who	0.86	0.37	1.32	0.17
When	0.69	0.50	1.08	0.09
Where	0.60	0.33	0.60	0.17

"What is history?" Students were asked to answer several questions in writing concerning their thoughts about what history is, what kinds of history they had learned on their own, and the importance of learning history for students their age. To date, responses to the first question, "In your own words, what is history?" have gone through preliminary analysis has been completed and it will be reported later when analysis is complete, although it looks as though the item may not have been sensitive to differences between student groups.

Students were asked to provide written responses to the question "In your own words, what is history?" The purpose of this survey was to determine how students think about history. Other empirical studies have examined how teachers and historians define history and how their definitions affect their students' understanding of history (Evans, 1988; Leinhardt, et al., 1994). Evaluators hoped to discover the nature of the typical elementary student's conception of history

and to see if using a survey to elicit a written definition would work as a means of exploring students' understanding of history.

The rubric used to code student responses was constructed in several stages. The same instrument was used with another TAH project that CEA was evaluating at the same time and the instrument and rubrics were developed using the same procedures for both projects, providing a larger number of student responses with which to derive meaningful codes. A historian, a social studies methods professor, and the *BHHII* project director were asked to generate examples of "ideal" and "typical" teacher and student responses to this question. Their examples were used to construct a rubric for teachers, which was then adapted for students. Two evaluators read samples of student responses to validate and refine the rubric to make it more appropriate for student responses. Samples of student responses were independently coded by two evaluators using the rubric; when disagreement arose, a consensus approach was used. Once evaluator agreement became consistent, the rubric was considered complete, and training materials were developed. A trained rater coded all student responses. During training, the rater met with evaluators and went over several small samples of student responses to further illustrate how to apply the rubric. The rater then independently coded small samples of student responses. Evaluators monitored the extent of agreement and made corrections and modifications in coding procedures where necessary. Complete instructions for rater training and the "What is history?" assessment form are included in the appendices of this report.

3.3 Classroom Observations

Members of the CEA evaluation team visited *BHHII* classrooms at all grade levels during the 2005-06 school year. A protocol for describing the classroom observations was developed during the original *BHH* project and during Year 2 of the *BHHII* Project. In order to systematically describe the classroom activities, an observation protocol was used that required observers to classify the types of activities that took place during the class, quantify the amount of time spent on each activity, make judgments as to the students' and teachers' level of engagement during the activities, and write a brief description of the main classroom activities. Table 10 shows the number of visits to each school and grade level.

Table 10. Number of BHH II Classroom* Observations by Grade and School

Grade	Total Number of *BHHII* Classrooms	School 1	School 2	School 3	Total Number of Visits
K	17	2	4	0	6
1	17	0	3	2	5
2	16	2	0	4	6
3	16	4	2	0	6
4	15	0	6	3	9
5	15	5	5	4	14
All Grades	96	13	20	13	46

*In all but four cases, the number of classrooms is the same as the number of teachers.

The numbers and locations of classroom observations were not randomized across the teacher participants or schools because of a number of factors including scheduling difficulties, distance to travel to schools, weather, and travel conditions. The evaluators made every attempt to observe as many different situations as possible, but the observations should not be considered to be exhaustive or representative, but primarily a convenience sample.

The main purpose of the observations was to determine the extent to which students in *BHHII* classrooms were actively engaged in learning history. When all the observations were complete, the evaluation team established a holistic rubric for assigning a score to each observation. The rubric employs a five-point scale based on several dimensions. The dimensions are:

- Student activity – extent to which students are actively learning history
- Student engagement – extent to which students appear engaged in learning history
- Stimulation for historical thought – extent to which the classroom situation (curriculum, materials, and teacher) supports student use of historical thinking skills
- Teacher engagement – extent to which teachers display enthusiasm and interest in teaching history

An additional dimension was observed; the extent to which the lesson observed was a history lesson and/or used the *BHH* curriculum. There was little variation among classrooms on this dimension because nearly all classrooms used the curriculum or history-based teacher modifications for the entire class period, therefore this dimension was not considered in the final holistic scoring of observations.

For scoring, the observation protocols were divided into two groups; Kindergarten through 2nd grade and 3rd grade through 5th grade. There were two reasons for grouping the observation data in this way. Using the larger groups provides more bases for looking across

classrooms rather than using the limited number at each grade, and the curricular formats and activities used in the 3rd through 5th grade are more similar to each other than they are to those used in the younger grades. Although the 2nd grade curriculum has elements of both the younger and older grade's curricular formats, the nature of the activities carried out in 2nd grade is more like the younger grades, so the 2nd grade observations were grouped with Kindergarten and 1st grade observations.

The lead evaluator read all the observation protocols and used the dimensions of the rubric to sort observation protocols into five levels, 1-5, where a score of 5 means:

- Students are *actively* learning history
- Students are *highly engaged* in learning history
- Students are *provided with multiple opportunities* to use and exhibit use of historical thinking skills
- Teachers *display high enthusiasm and interest* in history instruction

A score of 1 means:

- Students are *passive* or are not involved in learning history
- Students are *not engaged* in learning
- Students are *given few or no opportunities* to use and exhibit use of historical thinking skills
- Teachers *do not exhibit enthusiasm or interest* in history instruction

Scores of two to four mean that the ratings of at least some of the dimensions fall somewhere in-between those two extremes. In classrooms assigned mid-range holistic scores, it is possible that one or more dimension may be rated very high or very low, but the overall impression would preclude the classroom observation from earning the extreme scores. The complete observation protocol form is provided in the appendices to this report.

3.4 Teacher Focus Groups

Evaluation team members conducted focus groups with grade level teams in each of the six grades participating in the *BHHII* project. Two focus groups were conducted at each of the three schools. The groups were selected using a stratified random sample among the 18 possible groups, so that a group would be chosen at each grade, two would be chosen at each school, and at each school one group would be with a lower grade level (K-2nd grade teachers), and the other would be with an upper grade level group (3rd -5th grade teachers). Focus groups that comprised only grade level peers were chosen so that teachers could talk about the specific curriculum used at their grade level.

The same format was used and questions were asked at all grade levels. All groups occurred during May of the 2006, when nearly all the teachers had completed their *BHH* curriculum for the year. Evaluators emailed teachers and asked for available times to meet as a group, either before school, during lunch, or after school on a school-day. All groups chose to meet after school except one group that met with the evaluator during their lunch hour. Lunch or snacks were provided to all participants. Questions were emailed to all participants approximately one week prior to the group for their consideration. The groups lasted approximately 30-45 minutes. The questions concerned the benefits and problems for teachers associated with teaching the *BHH* curriculum, benefits and drawbacks for students, factors affecting their ability to successfully implement the curriculum (both positive and negative), thoughts on future implementation of the curriculum, ideas for improving professional development for use with the *BHH* curriculum, and general thoughts on teaching history in an elementary school setting.

Focus groups were recorded, transcribed, and summarized by members of the evaluation team. Full lists of questions, redacted transcriptions and summaries are included in the appendices to this report.

3.5 Student Group Interviews

During the 2005-06 school year, evaluators conducted small group interviews with students in 3rd to 5th grade, from three treatment schools and one comparison school. Three different evaluators conducted the group interviews. Groups typically consisted of five students from one classroom, at one grade level. Interviews took place in vacant classrooms or meeting rooms at the school and lasted approximately 30-45 minutes. Students within each group were selected by the teacher but the teacher was asked to select students of varying general ability levels. In most cases, the teachers told evaluators, without identifying which, that one of the students was of lower ability, two or three were medium ability level and one or two were of high ability level. Table 11 shows the number of groups conducted and the distribution of groups across schools, grades, and *BHH* curriculum units. Since only one or two groups were conducted at each school and grade level, these groups were not meant to create a complete picture, but to provide a snapshot of student behavior.

Table 11. Number of BHH II Student Group Interviews Conducted at Each Grade and School

	Grade 3*	Grade 4**	Grade 5***
Treatment School 1	2	2	2
Treatment School 2	0	2	2
Treatment School 3	2	2	2
Comparison School	1	1	1

*The topic for Grade 3 interviews was the Segregation unit
**The topic for Grade 4 interviews was the Great Depression unit
***The topic for Grade 5 interview was WWII Home Front unit, except at Treatment School 1, where the Native American unit was discussed.

The purpose of the interviews was to provide another look at students' knowledge of history content and their ability to use the skills taught in the *BHH* curriculum. We wanted students to have a chance to display their knowledge verbally rather than in the written format used for the other student assessments. In addition, the group format allowed students to expand and build on the responses of other students to provide an indication of the depth of student knowledge concerning a particular topic. Each interview was crafted to address the specific content of the unit that each group had recently studied and to include opportunities to demonstrate some of the skills used during the unit, most frequently, photo and/or poster analysis. Fifth grade students also used a series of photographs to help them create a narrative about a time in history. The *BHH* curriculum uses some form of photo analysis as a tool for developing skills for learning history at all grade levels.

Interviews also included questions concerning the students' opinions on learning history, likes and dislikes, and what they might want to learn about history. All groups began with an icebreaker question asking them, "If you could live at any time in history, when would you choose to live and why?"

Focus groups were tape recorded, transcribed, and summarized. Complete focus groups protocols, transcripts, and summaries are included in the appendices to this report.

4. Evaluation Questions

The general evaluation questions addressed in the Year 1 and Year 2 evaluation reports concerning the *BHHII* Project were:

- What is the impact of the professional development workshops on teachers in three different districts?
- Can the *BHH* curriculum be packaged to allow successful web-based dissemination of curricular materials and resources?
- What is the role of project staff in preparing teachers to teach the *BHH* curriculum?
- What is the role of teacher mentors in preparing teachers?
- In what ways can other institutions (e.g. *BHHII* partners IPTV, Hometown Perry Museum) contribute to the dissemination and implementation of history curriculum?
- What other kinds of activities occur to further the dissemination of the *BHH* curriculum?

During Year 2 and Year 3 the following evaluation questions were examined and are reported here:

- How many teachers implement the *BHHII* curriculum?
- In what ways are implementations of the *BHHII* curriculum different (if at all) between teachers and schools?
- How many students experience the *BHHII* curriculum?
- What kinds of skills and knowledge do students gain as a result of learning through the *BHHII* curriculum?
- How did the performance of comparison students differ from that of treatment students?
- How was the instruction provided by treatment teachers different from comparison teachers?
- How did the ability of students to think historically change as a result of the implementation of the BHH units?
- Do treatment students know more about historical content than comparison peers?
- In what ways is the *BHH* curriculum disseminated?

5. Evaluation Results

5.1 Teacher Outcomes

5.1.1 Implementation of History Curriculum

Participating treatment teachers and comparison teachers responded to three surveys concerning their implementations of history curriculum during the final year and a half of the *BHHII* project, a pilot administration during the Spring of the 2004-05 school year, and finalized surveys during the Fall and Spring of the 2005-06 school year. The surveys' scaled items served five main purposes; 1) to determine the teachers' self-described thoroughness of teaching different historical topics as part of the *BHH* curriculum, 2) to ascertain teachers' perceptions of their students' competence at performing skills or demonstrating content knowledge related to the *BHH* curriculum, 3) to gather teachers' opinions of the benefits and drawbacks of the curriculum for their students, 4) to determine teachers' perceptions of the utility of different pedagogical techniques and instructional practices for teaching history, and 5) to ask teachers' opinions on teaching history, interest in learning history, and self-efficacy concerning teaching history. The first set of surveys served primarily as a pilot for subsequent surveys and results are not reported here. In addition to the scaled items, a number of open-ended questions intended to divulge more about teachers' self-described thoroughness of implementation, other history topics and skills taught, the perception of other student competencies in historical content and skills, and other effects of teaching the *BHH* curriculum.

Although the same categories for items appeared on all surveys, twelve different versions of the survey (two at each grade level in grades K-5) reflected the content and focus of the twelve different *BHH* curricular units. The full surveys and results are included in the appendices to this report. Highlights of these results will be discussed (organized by grade and topic within survey section) in this section.

Thoroughness of Implementation

In each of the grade and topic specific surveys, treatment and comparison teachers rated the thoroughness of their implementations of each of the *BHH* curricular units through scaled items that roughly corresponded with individual activities designated in the curriculum. The thoroughness scale for these items ranged from "0" meaning "Not at all", to "5" meaning "Very thoroughly". In most cases, the item stem reflected the title of the activity used in the *BHH* curriculum. If the titles of the activities were not descriptive enough to be clearly understood as written, they were modified for the survey to make clear to both treatment and comparison

teachers exactly what content knowledge or skill the item was intended to address. If the activity title in the curriculum was not specific to particular content or a skill to be taught, such as "Final Project", these items were not used on the comparison teacher version of the survey. A summary of the teachers' self-described thoroughness of implementation across grades and *BHH* units is provided below, followed by details of these results at each grade and unit level. The full results for scaled items and summaries of open-ended item responses are found in the appendices to this report.

Summary of Thoroughness of Implementation across topics and grades

Teachers were asked, through a series of scaled items specific to each grade level and unit, to rate the thoroughness of their own implementation of specific activities that were part of the *BHH* curriculum. Across all grade levels and units, *BHH II* treatment teachers' rated their thoroughness of implementation of instruction on the topics and skills that are part of the *BHH* curriculum higher than their comparison teacher peers.

Table 12 shows mean results for all scaled items for Thoroughness of Implementation from the teacher surveys administered during the 2005-06 school year. Teacher ratings were averaged over all items within respective item clusters (i.e., Thoroughness, Perception of student competencies). For the treatment teachers, all but one of the mean ratings of the thoroughness of implementation was above 3.0, indicating that they thought that on average they had implemented the unit activities at least "moderately well" (4th grade teachers' mean rating of thoroughness of their implementations on *Progressive Era* topics was only slightly lower, at 2.95). For six of the twelve units, teachers' mean self-rating of how thoroughly they had implemented the units was above 3.5. On two topics, *WWII Home Front* (4.00) and *My 1st Grade History* (4.03), the treatment teachers' mean ratings were greater than or equal to 4.0 indicating that they thought they had implemented the activities at least "thoroughly". Therefore nearly all the treatment teachers rated their implementations of the *BHH* curriculum as being at least "moderately thorough" to "thorough".

In contrast, among the comparison group, comparison teachers' mean ratings of thoroughness of the implementation of instruction on the activities and topics included in all but three (out of 12) of the *BHH* curricular units, were between 1 and 2. On the scale used, a "1" rating means that they did not provide any instruction on the topic or skill at all, so thoroughness

Table 12. Grand Means for BHH II Treatment and Comparison Teachers by Item Type

Grade	Topic	Grand means for each item type cluster		
		Thoroughness of Implementation	Perception of Student Competencies	Perceived Benefit of Activities
	(n)	(k)	(k)	(k)
K	**History of Me**			
	Treatment (22)	3.58 (10)	3.53(19)	3.89(10)
	Comparison (6)	2.22(10)	2.29(19)	
	Children Long Ago			
	Treatment (22)	3.49(6)	3.66(12)	3.84(6)
	Comparison (6)	1.61(6)	1.78(12)	
1	**My 1st Grade History**			
	Treatment(25)	4.03(7)	3.83(18)	4.37(7)
	Comparison (8)	2.12(7)	2.61(18)	
	Community History			
	Treatment (25)	3.77(4)	3.87(10)	4.16(4)
	Comparison (8)	1.76(4)	2.45(10)	
2	**Immigration**			
	Treatment (16)	3.65(8)	3.67(17)	4.21(8)
	Comparison (7)	1.75(7)	1.57(17)	
	Environmental History			
	Treatment (21)	3.71(9)	3.53(26)	4.08(9)
	Comparison (7)	2.13(8)	1.96(26)	
3	**Industrialization**			
	Treatment (18)	3.24(9)	3.22(21)	3.29(9)
	Comparison (6)	1.87(9)	2.20(21)	
	Segregation			
	Treatment (18)	3.56(8)	3.30(19)	4.15(8)
	Comparison (6)	2.13(7)	1.96(19)	
4	**Great Depression**			
	Treatment (17)	3.41(9)	3.33(18)	3.72(9)
	Comparison (8)	1.90(8)	1.70(18)	
	Progressive Era			
	Treatment (14)	2.95(8)	3.30(23)	3.72(8)
	Comparison (7)	1.67(6)	1.94(23)	
5	**Native American History**			
	Treatment (15)	3.41(9)	3.22(20)	3.27(9)
	Comparison (5)	1.38(9)	1.65(20)	
	WWII Home Front			
	Treatment (14)	4.00(10)	3.71(17)	4.31(10)
	Comparison (4)	1.53(9)	1.65(17)	

ratings between 1 and 2, signify that the item was covered as part of classroom instruction somewhere between "Not at all" and "Only slightly". The low ratings seen nearly across the board among comparison teachers (the lowest treatment teacher mean rating is still nearly one whole point above the highest comparison teacher mean rating and these extremes are not on the

same topic), indicate that for the most part, the topics taught as part of the *BHHII* project are seldom or never taught in comparison schools. There were only four topics in which the comparison teachers' mean thoroughness rating exceeded 2.0 for comparison teachers; these occurred on the *History of Me (2.22), My 1st Grade History* (2.12), *Environmental History* (2.13) and *Segregation* (2.13) topics.

When talking with treatment teachers, these three topics were often mentioned as ones that they had also "taught before", although often in another format. For example, kindergarten teachers often said that the *History of Me* unit was similar to things they had done in previous years to help the teacher and students get to know each other (e.g. Star Student or Student of the Week), but that the *BHH* curriculum brought a new perspective by making students' personal stories and "show and tell" become their history and adding new vocabulary like "artifacts". *Environmental History* appears to be a good fit with a common 2nd grade emphasis on the environment, but again with a new emphasis on history and past efforts to protect the environment.

Similarly, third graders often receive instruction concerning slavery and Martin Luther King, Jr. but teachers felt that adding the aspects on Jim Crow laws, other segregation and civil rights era history, and the role of Constitutional Amendments as setting the *BHH* curriculum apart from what they had taught in the past and in responses to specific aspects of thoroughness of coverage by comparison teachers, that differentiated treatment and comparison instruction on the same topic area. For example, comparison teachers mean rating on the thoroughness of their coverage of the topic *The 13th Amendment* was 1.33 (compared to 3.33 for treatment teachers) *Ways that African American people resisted segregation* was 1.83 (compared to 3.50 for treatment teachers) and for the topic, *African American people during the segregation years, 1865-1950,* the mean rating was also 1.83 (compared to 3.78 for treatment teachers). The greatest differences between treatment and comparison teachers mean ratings on the thoroughness of their implementations on the topics specified were in the *Native American History* (5th grade), *WWII Home Front* (5th grade), *Community History* (1st grade), *My 1st Grade History* (1st grade), and *Immigration* (2nd grade).

In addition, student competencies for treatment/comparison classrooms were somewhat discrepant for those units, suggesting that even though comparison teachers covered some of the same content, comparison teachers perceived their students as being less competent at related skills than treatment students were perceived.

In general, third and fourth grade treatment teachers rated their curriculum implementations as less thorough than did treatment teachers of other grades. In particular,

according to the mean teacher ratings of thoroughness, three units were implemented less thoroughly than the others, the 4th grade *Progressive Era* unit (with a mean thoroughness rating of 2.95), the 3rd grade *Industrialization* unit (3.24), and the 5th grade *Native American History* unit (3.41). During focus group interviews and in casual conversation with teachers, the same three units were also almost uniformly named as less popular with teachers, saying that they were less comfortable with the topic area or they deemed the lessons either too difficult or not as interesting to students as the other units. One 5th grade teacher said, "When we did the *Native American* unit, it was very difficult for us as teachers to get them involved and to get them excited about it. So it was kind of a hard unit." A 3rd grade teacher commented, "In terms of background knowledge, I felt better with this unit than I ever did with the *Industrial Revolution*. I just felt like I was kind of out there with that one and I was learning it all brand new. But I was kind of lost with the *Industrial Revolution*." Sometimes it was just part of the unit that the teacher found less useful, such as the comments 4th grade teachers made concerning the *Progressive Era* unit, "… there was an initial part to the Progressive Era that we questioned if it needed to be in there. It had to do with the…"(another teacher finished the sentence with) "…business and economy."

The same three units also received the lowest overall teacher ratings of perceived benefit for students, with mean ratings of 3.27 for the 5th grade *Native American* unit, 3.29 for the 3rd grade *Industrialization* unit, and 3.72 for the 4th grade *Progressive Era* unit. (A mean rating of 3.72 was also given by the 4th grade teachers to the *Great Depression* unit, but this lower rating was not seen in their other comments or survey responses.)

At the other grades, teachers sometimes expressed more interest in or preference for one unit than the other, but typically not in as negative characterizations as the comments from the teachers given above, usually more in the vein of they especially liked one unit. Teachers were encouraged to continue to share their ideas with the Project Director and revisions of those three units are in process. However, it should be noted that even for these comparatively less popular units, treatment teachers reported more through implementation of the topic matter than did their comparison peers.

In addition to the scaled items concerning implementation thoroughness, teachers were asked whether they were able to implement the *BHH curriculum unit* to the extent that they hoped. Among the treatment teachers, 115 of the 119 total *BHHII* teachers (97%) replied to this item with 46% of those responding saying simply "Yes", they had implemented the curriculum to their satisfaction and a few mentioning that the students were interested and engaged and that the curriculum fit well with their pre-existing curricula. One teacher said, "Yes, definitely, the kids loved it" and another teacher said, "Yes, enjoyed teaching it."

The other 54% of the treatment teachers said they had not implemented the curriculum to the extent that they had hoped with the main reasons given for this being: general time constraints (44% of those who said "No"), school more focused on other content areas, NCLB, or assessments (26%), not enough prep time or unfamiliarity with lesson or topic (11%) and curriculum too difficult or other unit specific reasons (10%). A typical comment written by one teacher was, "Time constraints (end of year testing, etc.) left me scrambling to finish by the end of the year, so there were some books and lessons that weren't studied in as much depth as I would have wanted to do." Another teacher said, "It was difficult to allow time for the whole unit. We have so many *No Child Left Behind* assessments to take care of. It made it impossible to do all of the *BHH* lessons." All of the comments concerning the difficulty level of the units as a hindrance to thorough implementation occurred in the fourth grade, in reference specifically to the Progressive Era unit. One teacher said, "No, because of time, but also because it was hard topics for them to discuss [in the Progressive Era unit] in comparison to Great Depression."

The comparison teachers were asked if they were able to teach *history* as thoroughly as they would have liked. Nearly all (38 out of 39) of the comparison teachers responded for a response rate of 97%. Their responses were also split, with 21% indicating that they felt they had taught history to the extent that they would like (again most of these teachers simply said "yes" without expanding on that) and 66% indicating that they had not. (The additional 13% gave other answers such as that they taught social studies more than history without indicating whether they were satisfied by that situation.) Among those who said they did not feel they had taught enough history, about half said because of time constraints and half said because other content areas are given priority over history. One comparison teacher said, "Not at all. With the *No Child Left Behind* legislation, we are focusing a lot of our time and energy on language arts skills and math. Social studies and history are plugged in wherever we can find a little time for it." Another teacher said, "No, teaching reading and math are much more important than teaching history." A few said they lack resources and funding to teach history.

Grade Level Details Concerning Thoroughness of Implementation

Kindergarten

History of Me

Teachers responded to ten items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit, *The History of Me.* Treatment teachers indicated that they implemented eight of the ten topics at least moderately well (a mean score 3.0 or above) and for five of the ten items the mean rating of their implementation was at least 3.5. The mean level of

implementation across the ten topics was 3.58. The treatment mean exceeded the comparison mean for all items, some with non-overlapping or nearly non-overlapping distributions of scores.

For comparison teachers, eight of the ten topics surveyed had a mean of less than 3.0 and five of the topics had a mean less than 2.0. The mean across all ten topics was 2.22.

Children Long Ago

Teachers responded to six items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Children Long Ago*. Treatment teachers indicated that they implemented all of the six topics at least moderately well (a mean score 3.0 or above) and for three of the six items the mean rating of their implementation was at least 3.5. The mean level of implementation across the six topics was 3.49. The treatment mean exceeded the comparison mean for all items, some with non-overlapping or nearly non-overlapping score distributions.

For comparison teachers, all of the six topics surveyed had a mean of less than 2.0. The mean across all eight topics was 1.61. It is important to note that for this scale, a rating of "1" means that the topic was not covered at all.

First Grade

My First Grade History

Teachers responded to seven items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *My 1st Grade History*. Treatment teachers indicated that they implemented all of the seven topics at least moderately well (a mean score 3.0 or above) and for five of the seven items the mean rating of their implementation was at least 4.0. The mean level of implementation across the seven topics was 4.03. The treatment mean exceeded the comparison mean for all items, some with non-overlapping and nearly non-overlapping distributions of scores.

For comparison teachers, six of the seven topics surveyed had a mean of less than 3.0. The mean across all seven topics was 2.12. It is important to note that for this scale, a rating of "1" means that the topic was not covered at all.

Community History

Teachers responded to four items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Community History*. Treatment teachers indicated that they implemented all of the four topics at least moderately well (a mean score 3.0 or above) and for two of the four items the mean rating of their implementation was at least 3.5. The mean level of implementation across the four topics was 3.77. The treatment mean exceeded the comparison mean for all items, some with nearly non-overlapping distributions of scores.

For comparison teachers, all of the four topics surveyed had a mean of less than 3.0 and for three of the four topics, the mean level of thoroughness of implementation was less than 2.0. The mean across all eight topics was 1.76. It is important to note that for this scale, a rating of "1" means that the topic was not covered at all.

Second Grade

Immigration

Teachers responded to eight items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Immigration*. Treatment teachers indicated that they implemented all of the eight topics at least moderately well (a mean score 3.0 or above) and for five of the eight items, the mean rating of their implementation was at least 3.5. The mean level of implementation across the eight topics was 3.65. The treatment mean exceeded the comparison mean for all items, some with non-overlapping and near non-overlapping score distributions.

For comparison teachers, all except one of the seven topics surveyed ("Exchange of Cultures Festival" was not on the comparison survey) had a mean of less than 3.0, with six of the seven topics having a mean of less than 2.0. The mean across all seven topics was 1.75.

Environmental history

Teachers responded to nine items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Environmental History*. Treatment teachers indicated that they implemented all of the nine topics at least moderately well (a mean score 3.0 or above) and for seven of the nine items the mean rating of their implementation was at least 3.5. The mean level of implementation across the nine topics was 3.71. The treatment mean exceeded the comparison mean for all items, some with nearly non-overlapping distributions of scores.

For comparison teachers, ratings on all except one of the eight topics surveyed ("Super Heroes for the Environment: Conclusion and Review" was not on the comparison survey) had a mean of less than 3.0. The mean across all eight topics was 2.13. The highest mean for comparison teachers occurred on the item, "How farming methods have changed over time" (3.14).

Third Grade

Industrialization

Teachers responded to nine items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Industrialization*. Treatment teachers indicated that they

implemented five of the nine topics at least moderately well (a mean score 3.0 or above). The mean level of implementation across the nine topics was 3.24. The topics that scored below a 3.0 were, "The History of Corporations – Finding the money to build big business" (mean=2.53), "Industrialization patterns – how cities grew up around industry" (mean=2.78), "How to create and use a mind map to synthesize different things students have learned about history" (mean=2.57), and "The pros and cons of industrialization at different points in history" (mean=2.87). The treatment mean exceeded the comparison mean for all items.

For comparison teachers, ratings of all of the nine topics surveyed had a mean of less than 3.0 and seven of the nine topics had a mean of less than 2.0. The mean across all nine topics was 1.87. It is important to note that on this scale, a score of "1" means the topic is not covered at all.

Segregation

Teachers responded to eight items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Segregation*. Treatment teachers indicated that they implemented seven of the eight topics at least moderately well (a mean score 3.0 or above). The mean level of implementation across the nine topics was 3.56. The topic that scored below a 3.0 was, "The US Constitution" (mean=2.67). This item was the only item where the comparison mean (3.00) exceeded the treatment mean.

For comparison teachers, all except one of the seven topics surveyed ("Review and Conclusion" was not on the comparison survey) had a mean of less than 3.0 and four of the seven topics had a mean of less than 2.0. The mean across all seven topics was 2.13. The highest means for comparison teachers occurred on the items, "The US Constitution" (3.0), "Prejudice and tolerance" (2.67) and "The history of prejudice against African Americans (2.50). In open-ended items comparison teachers said that they taught about the Martin Luther King, Jr. and slavery during their African American history units and, as of 2004, all federally-funded schools are required by law to teach about the Constitution.

Fourth Grade

The Progressive Era

Teachers responded to eight items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on the *Progressive Era*. Treatment teachers indicated that they implemented two of the eight topics at least moderately well (a mean score 3.0 or above). The mean level of implementation across the nine topics was 2.95. The general level of thoroughness of implementation of these activities was lower than for other units, but there was

some variation between topics on the thoroughness of teacher implementation with two topics rated as being fairly thoroughly covered, "History of Child Labor" (mean=3.93) and "Tenement Housing" (mean=3.71) and two topics rated as just higher than "only slightly covered", "History of Capitalism" (mean=2.43) and "Modern Day Progressives – Civic Action" (mean=2.45). In all cases, treatment means were higher than the comparison means for the same items.

For comparison teachers, all of the six topics surveyed (the activities "Final Activity: Great Depression Autobiography" and "Overview of Industrial Society – Review of immigration, industrialization, and corporations" were not on the comparison survey) had a mean of less than 3.0 and four of the six topics had a mean of less than 2.0. The mean across all six topics was 1.67, meaning they taught the topic "only slightly" or less.

The Great Depression

Teachers responded to nine items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on the *Great Depression*. Treatment teachers indicated that they implemented seven of the nine topics at least moderately well (a mean score 3.0 or above). The mean level of implementation across the nine topics was 3.4. The topics that scored below a 3.0 were, "How to create and use a mind map to synthesize different things students have learned about history" (mean=2.71) and "Final Activity: Great Depression Autobiography" (mean=2.18). All of the treatment means were higher than the comparison means for the same items.

For comparison teachers, all of the eight topics surveyed (the Final Activity: Great Depression Autobiography was not on the comparison survey) had a mean of less than 3.0 and five of the eight topics had a mean of less than 2.0. The mean across all eight topics was 1.9, meaning they taught the topic "only slightly" or less. The highest mean for comparison teachers (2.88) occurred on the item "The Dust Bowl." In open-ended items, comparison teachers said that they taught about the Dust Bowl during their regions unit on the Southwest. It is important to note that on the scale used for this task the option "1" actually stands for "Not at all", meaning that comparison teachers taught very little about the Great Depression.

Fifth Grade

Native Americans

Teachers responded to nine items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on *Native American History*. Treatment teachers indicated that they implemented seven of the nine topics at least moderately well (a mean score 3.0 or above). The mean level of implementation across the nine topics was 3.41. The topics that scored below

a 3.0 were, "Locations of Reservations" (mean=2.73) and "Problems for contemporary Native Americans" (mean=2.60). Means on all items were higher for treatment teachers than for comparison teachers.

There were only five fifth-grade comparison teacher respondents, so conclusions made from their responses may be limited, however, the magnitude of the differences between the two groups was large, with score distributions non-overlapping or nearly non-overlapping. For comparison teachers, none of the nine topics surveyed scored higher than "only slightly" (a mean score of 2.0) and the mean across all nine topics was 1.38. It is important to note that on the scale used for this task the option "1" actually stands for "Not at all" meaning that comparison teachers taught very little about Native American History.

WWII – The Home Front

Teachers responded to ten items on a 5-point scale concerning the thoroughness of their implementation of the *BHH* unit on the *WWII Home Front*. Treatment teachers indicated that they implemented nine of the ten topics at least moderately well (a mean score 3.0 or above). The mean level of implementation across the ten topics was 4.0. The topic that scored below a 3.0 was "What it was like on the home front when the war was over" and the mean score for treatment teachers on this topic was a 2.9.

There were only four fifth grade comparison teacher respondents, so conclusions made from their responses may be limited, however, the magnitude of the differences between the two groups was large. In addition, comparison teachers from one school are departmentalized, meaning that they teach only social studies/history so may be more comfortable with and teach more history than their self-contained classroom peers. For comparison teachers, none of the nine topics surveyed (the "Final Project" activity was not on the comparison survey) scored higher than an "only slightly" (a mean score of 2.0) and the mean across all nine topics was 1.5. It is important to note that on the scale used for this task the option "1" actually stands for "Not at all" meaning that comparison teachers taught very little about the WWII home front.

5.1.2 Attitudes toward teaching history and other teacher outcomes

Teacher Self-Efficacy in Teaching History (and other topics)

The fourth part of the survey asked teachers to rate their level of agreement or disagreement with five items on a six-point Likert type scale concerning their comfort in teaching history, interest in learning more history, perception of student attitudes toward learning history, importance of teaching history, and comfort in using primary sources in their classroom.

Kindergarten

History of Me

These items were not included on this survey.

Children Long Ago

Responses to these items were overwhelmingly positive for both groups with nearly all teachers at least mildly agreeing with each item and most agreeing or strongly agreeing. On four of the five items, the treatment group mean slightly exceeded the comparison group mean. Across the five items, the treatment group mean was 5.1 and the comparison group mean was 4.5. The one item where the comparison group mean (5.17) was greater than the treatment group mean (4.71) was the item, "I am interested in learning more history." The largest differences were on the item "I feel comfortable teaching history" (treatment mean=5.18, comparison mean=4.00) and "I feel comfortable using primary sources in my history instruction" (treatment mean=5.19, comparison mean=4.00), but both groups were quite positive about all the statements.

First Grade

My First Grade History

These items were not included on this survey.

Community History

Responses to these items were overwhelmingly positive for the treatment group with nearly all teachers at least mildly agreeing with each item and most agreeing or strongly agreeing. Responses were also generally positive for comparison teachers. On each of the five items, the treatment group mean slightly exceeded the comparison group mean. Treatment teachers appeared to be slightly more comfortable teaching history (treatment mean=5.08, comparison mean=4.38), find teaching history to be more important to them (treatment mean=5.12, comparison mean=4.5) and feel more comfortable using primary sources to teach history (treatment mean=5.27, comparison mean=4.00). Across the five items, the treatment group mean was 5.1 and the comparison group mean was 4.5.

Second Grade

Immigration

These items were not included on this survey.

Environmental history

Responses to these items were overwhelmingly positive for both groups with nearly all teachers at least mildly agreeing with each item and most agreeing or strongly agreeing. On each of the five items, the treatment group mean slightly exceeded the comparison group mean. Across the five items, the treatment group mean was 5.2 and the comparison group mean was 4.7.

Third Grade

Industrialization

These items were not included on this survey.

Segregation

Responses to these items were overwhelmingly positive for both groups with nearly all teachers at least mildly agreeing with each item and most agreeing or strongly agreeing. On four of the five items, the treatment group mean exceeded the comparison group mean. Across the five items, the treatment group mean was 5.0 and the comparison group mean was 4.7.

Fourth Grade

The Progressive Era

These items were not included on this survey.

The Great Depression

These items were inadvertently left off the surveys sent to all but one treatment school and one comparison school. From this limited sample, responses to these items were overwhelmingly positive for both groups with all teachers at least mildly agreeing with each item with the exception of two items where on each item, one comparison teacher mildly disagreed. On all items except one, the treatment teacher mean agreement exceeded the comparison teacher group mean, although differences between groups were small and not practically significant.

Fifth Grade

Native Americans

These items were not included on this survey.

WWII – The Home Front

Again there were only four comparison teachers responding to these items and one comparison teacher chose the "No Opinion" option, so differences between the two groups may not be important. Responses to these items were overwhelmingly positive for both groups with all teachers at least mildly agreeing with each item with the exception of two items where on each item, one treatment teacher mildly disagreed. In this section, on all items except one, the

comparison teacher mean agreement exceeded the treatment teacher group mean. Two of the four responding comparison teachers teach only social studies/history, so their comfort and interest in teaching and learning history is perhaps to be expected. The item where the treatment group mean was greater than the comparison group mean, was the item, "I feel comfortable using primary sources in my history instruction". For that item the treatment mean was 5.2 and the comparison mean was 4.0. *BHH* professional development and curriculum stress the use of primary sources in instruction.

Use of Pedagogical and Instructional Practices

The fourth part of the survey asked teachers to respond to six items on a five-point scale concerning the usefulness of various pedagogical and instructional practices for teaching history that are suggested for use in teaching the *BHH* curriculum. The same items were used on all twelve surveys. The teachers were asked to answer the items in reference to *teaching history*, but without specific reference to the particular unit they had most recently completed, therefore teachers responded to the exact same items twice during the same school year and they serve as a reliability estimate for scaled portions. The five-point scale used for these items ranged from "0" meaning the teacher found that specified practice "Not at all useful" to "5" meaning the teacher found the practice "Very useful."

Kindergarten

History of Me

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, treatment teachers found the six pedagogical or instructional practices to be at least moderately useful, rating all six of the items at least a 3.0. The treatment teacher mean across the six items was 3.83. The treatment teacher mean exceeded the comparison teacher mean for all six of the practices. (Comparison teachers and treatment teachers were nearly identical on the item "Drawing connections between new units and units encountered previously" with treatment teacher mean=3.86 and comparison teacher mean=3.75.) Comparison teachers found the pedagogical and instructional practices somewhat less useful with a mean across the six items of 2.41. The largest differences were found in "Use of primary source documents (treatment mean=3.74, comparison mean=1.50) and in "Student Presentations" (treatment mean=4.50, comparison mean=1.00). The "History of Me" unit centers around children telling

about their lives through the use of photographs and artifacts of their childhood, providing a combination of student presentations with use of primary sources.

Children Long Ago

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, treatment teachers found the six pedagogical or instructional practices to be at least moderately useful, rating five of the six items at least a 3.0. The treatment teacher mean across the six items was 3.79. The treatment teacher mean exceeded the comparison teacher mean for five of the six practices. (Comparison teachers and treatment teachers were nearly identical on the item "Drawing connections between new units and units encountered previously" with treatment teacher mean=3.95 and comparison teacher mean=4.00.) Comparison teachers found the pedagogical and instructional practices somewhat less useful with a mean across the six items of 3.07. The largest differences were found in "Use of primary source documents (treatment mean=3.82, comparison mean=2.60) and in "Student Presentations" (treatment mean=2.88, comparison mean=1.33).

First Grade

My First Grade History

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0 (the comparison teachers rated "student presentations" slightly below that, a mean of 2.63). The treatment teacher mean across the six items was 4.21. The treatment teacher mean exceeded the comparison teacher mean for all six practices. Comparison teachers found the pedagogical and instructional practices slightly less useful with a mean across the six items of 3.29. The largest differences were found in "Use of primary source documents (treatment mean=4.50, comparison mean=3.33) and in "Discovery learning (treatment mean=4.33, comparison mean=3.13).

Community History

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, treatment teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0. The treatment teacher mean across the six items was 3.99. The treatment teacher mean exceeded the comparison teacher

mean for all six practices. Comparison teachers found the pedagogical and instructional practices less useful with a mean across the six items of 3.06. The largest differences were found in "Use of primary source documents (treatment mean=4.36, comparison mean=2.71) and in "Discovery learning (treatment mean=3.83, comparison mean=2.50).

Second Grade

Immigration

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, treatment teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0. The treatment mean across all items was 3.88 and for comparison teachers, the mean was 2.77. The treatment mean exceeded the comparison mean for all items. The greatest differences between the two groups were seen on usefulness of group work (treatment mean=4.25, comparison mean=2.86) and usefulness of student presentations (treatment mean=3.77, comparison mean=2.00).

Environmental history

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately, useful rating each of the items at least a 3.0, with the exception of "student presentations" on which the comparison teachers mean was 2.42. There were only small differences between the two groups except treatment teachers were more likely to find "group work" useful (mean=4.33) than comparison teachers (mean=3.57) and treatment teachers also found student presentations more useful (a mean of 3.89 for treatment teachers and a mean of 2.42 for comparison teachers). The treatment mean across all items was 4.05 and for comparison teachers, the mean was 3.58.

Third Grade

Industrialization

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, treatment teachers found the most of the pedagogical or instructional practices to be at least moderately useful, rating all but one of the items at least a 3.0. The treatment mean across all items was 3.53 and for comparison teachers, the mean was 3.53. The

only item receiving a mean score from treatment teachers below a 3.0 was "Student presentations" (2.91). Comparison teachers also generally rated the pedagogical and instructional activities as useful, with the mean score for all but two of the items greater than 3.0. They also rated "student presentations" lower (mean=2.80) and also rated "use of primary source documents" lowest with a mean of 2.40. The treatment teachers found primary sources useful with a mean of "4.0 on that item. The comparison teacher mean rating was slightly greater than the treatment teacher mean rating for usefulness of "guided inquiry" (comparison mean=4.0, treatment mean=3.53) and for "Drawing connections between new units and units encountered previously" (comparison mean=4.20, treatment mean=3.50), although the means for those items were all quite high.

Segregation

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0. The treatment mean across all items was 3.69 and for comparison teachers, the mean was 3.76. There were only small differences between the two groups, in both directions. The largest differences appeared to be that the treatment teachers found using group work more useful (a mean of 4.28 for treatment teachers and a mean of 3.80 for comparison teachers). On the other hand, comparison teachers found using primary source documents more useful (a mean of 4.20 for comparison teachers and 3.71 for treatment teachers).

Fourth Grade

The Progressive Era

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0. On five of the six items, comparison teachers rated the practices as being more useful than did treatment teachers. The only one which treatment teachers rated as more useful was using primary sources (a mean of 4.31 for treatment teachers and a mean of 3.33 for comparison teachers). The treatment mean across all items was 3.79 and for comparison teachers, the mean was 4.15.

Comparison teachers were more likely to use guided inquiry (comparison mean=4.43, treatment mean=3.86), group work (comparison mean=4.43, treatment mean=3.79), student presentations (comparison mean=4.29, treatment mean=3.45), and found making connections between history units and other previously taught units to be more useful (comparison mean=4.71, treatment mean=3.62).

The Great Depression

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately, useful rating each of the items at least a 3.0. Differences between groups on individual items were small and the treatment and comparison means across all items were the same, 3.76. The largest differences were that comparison teachers found making connections between history units and other previously taught units to be more useful (a mean of 4.38 for comparison teachers and 3.65 for treatment teachers) and treatment teachers found using primary sources more useful (a mean of 4.15 for treatment teachers and a mean of 3.67 for comparison teachers).

Fifth Grade

Native Americans

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. Comparisons between the treatment and comparison teachers are difficult for some of these items because often one or more of the teachers chose the "No response" option, further limiting the number of teacher responses. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0. The treatment across all items was 3.61 and for comparison teachers, the mean was 3.82. There were very few differences between the comparison and treatment teachers on these items with comparison teachers rating some of the practices more useful than their treatment peers.

On this survey (as opposed to the 5th grade teacher survey on the WWII unit) comparison teachers found using primary source documents more useful than did treatment teachers (comparison mean=3.67, treatment mean=3.14). Treatment teachers said in open-ended questions that the documents associated with this unit were too difficult for students to understand and that appears to have affected their general response to this item. Comparison teachers also

found making connections between history units and other previously taught units to be more useful (a mean of 4.25 for comparison teachers and 3.38 for treatment teachers).

WWII – The Home Front

Teachers responded to six items on a five-point scale concerning their opinions of the usefulness of various pedagogical and instructional practices for teaching history in their classrooms. Comparisons between the treatment and comparison teachers are difficult for some of these items because often one or more of the comparison teachers chose the "No response" option, further limiting the number of comparison teacher responses. In general, both treatment and comparison teachers found the six pedagogical or instructional practices to be at least moderately useful, rating each of the items at least a 3.0. The treatment across all items was 4.02 and for comparison teachers, the mean was 3.90. The largest differences appeared to be that the treatment teachers found using primary sources more useful (a mean of 4.4 for treatment teachers and a mean of 3.3 for comparison teachers) and were more apt to find guided inquiry useful (a mean of 4.1 for treatment teachers and 3.0 for comparison teachers. On the other hand, comparison teachers found making connections between history units and other previously taught units to be more useful (a mean of 5.0 for comparison teachers and 3.9 for treatment teachers).

Summary of Usefulness of Pedagogical and Instructional Practices

Table 13 shows the grand means across all items on each survey concerning the use of certain pedagogical and instructional practices that are suggested for use with the *BHH* curriculum. Pooling the means for the six practices somewhat obscures some of the subtleties in usage that were found at particular grade levels (discussed above), but is interesting to demonstrate the teachers' perceptions of the general usefulness of the practices advocated by the *BHH* curriculum.

Table 13. Grand Means for All Teachers on Pedagogical and Instructional Practices Items

Grade	Topic	(n)	Pedagogical and instructional practices (k=6)
K	**History of Me**		
		Treatment (22)	3.83
		Comparison (6)	2.41
	Children Long Ago		
		Treatment (22)	3.79
		Comparison (6)	3.07
1	**My 1st Grade History**		
		Treatment(25)	4.21
		Comparison (8)	3.29
	Community History		
		Treatment (25)	3.99
		Comparison (8)	3.06

2	**Immigration**	
	Treatment (16)	3.88
	Comparison (7)	2.77
	Environmental History	
	Treatment (21)	4.05
	Comparison (7)	3.58
3	**Industrialization**	
	Treatment (18)	3.53
	Comparison (6)	3.33
	Segregation	
	Treatment (18)	3.69
	Comparison (6)	3.76
4	**Great Depression**	
	Treatment (17)	3.76
	Comparison (8)	3.76
	Progressive Era	
	Treatment (14)	3.79
	Comparison (7)	4.15
5	**Native American History**	
	Treatment (15)	3.61
	Comparison (5)	3.82
	WWII Home Front	
	Treatment (14)	4.02
	Comparison (4)	3.90*

*Only one comparison teacher answered the first item in this set, so that item was omitted and the mean of the other five items is reported here.

For teachers who teach kindergarten through 3rd grade, the treatment teachers were apt to find the specified practices more useful than their comparison peers. Practices that were particularly found to be more useful among treatment teachers than comparison teachers in grades K-3 were: use of primary sources, student presentations, and group work. All of the practices specified tend to be more commonly used in upper grade levels than they are in the lower grades. In 3rd through 5th grade, treatment and comparison teachers were similar and quite high in their rating of usefulness of the practices.

In addition to the scaled items concerning teachers' use of specific pedagogical and instructional practices, teachers were also asked to list the primary sources they used for teaching history and describe the ways in which they were used. A summary of the teachers' responses to this item can be found in the appendices to this report. Of the treatment teachers, 94 of the 115 teachers responded to this item for a response rate of 82%. Of these responses, the most common primary sources listed were (with response frequency): photographs (52), artifacts (20), letters (8), newspapers (4), copies of documents (found on internet) (4), videos from another time (4), journals (3), maps (3), WWII and runaway slave posters (3), and ration books (3). They said that these sources were used for photo analysis (20), document analysis (8), sorting photos by age (6), and making timelines (5). Five teachers in the treatment group said that they did not use any

primary sources and some gave responses that were unclear, such as "books" and a few listed things that are not primary sources, such as "easy-read" books and "*Little House on the Prairie* video."

Among the comparison group teachers 35 of the 38 teachers supplied a response to this item for a response rate of 92%. Of those responding, six comparison teachers said they did not use primary source. In the comparison teacher group, there appeared to be some differences between teachers who taught the lower and higher grade levels. Kindergarten through 2nd grade comparison teachers (the 14 of the 20 teachers from this group who replied to this item), supplied responses that were actually not primary sources, including picture books, globes, teaching materials, textbooks, library books, literature, videos and worksheets. It's possible that they were unfamiliar enough with the use of primary sources to the point where they misinterpreted the question to be "What are the *primary* (meaning main) sources you use when teaching history?"

The same thing occurred to a lesser extent with the 3rd through 5th grade comparison teachers, with 13 of the 18 teachers who responded saying that they used textbooks, although some of these teachers (6) indicated that they were referring to the primary sources that are included in their textbooks, such as journal entries, the Declaration of Independence, and newspaper articles. Only a few comparison teachers offered other specific examples of primary sources they had used such as photographs of local historical sites, journal excerpts, and legal documents. The comparison teachers tended to give fairly non-specific details as to how they were used saying things like they "read", "look at", or "discuss" them.

Other information concerning the teachers' experiences teaching history

Impact of BHHII on teaching of other history/social studies topics

Treatment teachers were asked two other questions about how their implementation of the *BHH* curriculum has affected their teaching. When asked how using the *BHH* curriculum has affected their teaching of other history or social studies, 106 of the 115 treatment teachers responded for a response rate of 92%. Many teachers described more than one way that their teaching had been affected by use of the *BHH* curriculum. Most responses fell into four categories, 1) change in the classroom experience, 2) change in quantity of history taught, 3) change in the teachers' attitude toward and/knowledge about history, and 4) teacher perception of change in their students' interest in learning history.

Responses concerning how using the curriculum affected their classroom experience included (frequency of mention in parentheses): increased use of primary sources (21), *BHH* provided a better, more focused structure for teaching history (12), made more cross-curricular

connections (7), and learned new strategies for teaching (5). One 2nd grade teacher commented, "It has encouraged me to use primary sources more often and has encouraged me to do more research on these topics. I've learned new things too. It also encourages me to have the students learn about more individuals who have influenced changes in our country."

Eleven teachers said that BHH influenced the quantity of history that they teach with six teachers saying that they teach more than they did before and five saying that they did not teach history at all before and now they do. One first grade teacher said, "We had not taught history in the past – now I realize that first graders can come to some understanding of history."

There were 37 comments by teachers concerning changes in their attitudes about or interest in teaching history with teachers saying that they are: more excited about learning history (14), have learned more history (9), have more fun teaching history now (7), it's easier to teach now (4), value teaching history more now (2), and have created their own *BHH* style unit (1). One third grade teacher said, "I have learned with the children and it has made history more enjoyable for me."

Twenty seven teachers commented on the impact of teaching history for their students' attitudes concerning history. Some said that kids like history more using the *BHH* curriculum (17), that it is more relevant to kids (6), and that they realize kids can learn more about history than they thought before (4). A fourth grade teacher said, "Don't underestimate interest level of fourth graders; if history is presented well students respond well."

Impact on teaching other subjects

Teachers were asked how participating in the *BHH* project had affected their teaching of other subjects. A total of 93 of the 115 teachers responded to this question for a response rate of 81%. The most frequently mentioned change in their teaching in general, mentioned by 26 teachers, was that they are more likely to make connections across the curriculum. One teacher said, "I find myself referring back to the unit during other discussions. I've also had students 'connect' to some of the vocabulary words presented during the unit as they run across them in other areas."

About one-fourth of the teachers (23) also said they are able to use the books they use for *BHH* in their shared reading or theme based units. One teacher commented "I incorporate more nonfiction into other subjects and use a variety of strategies to grasp concepts."

Nearly a third of the teachers who responded (30) mentioned that they've changed some of their strategies for teaching other subjects including:

- Using more discovery learning, inquiry, asking questions

- Using more primary sources
- Analyze photos in other areas
- Have students talk with each other about their learning
- Use more group learning situations
- Have students write more
- Use more hands-on activities

One teacher said, "I've added more discovery and group learning due in part to *BHH* and in part to classes taken through a masters program."

A few teachers said that teaching *BHH* has negatively impacted their teaching of other subjects because they had to take time from teaching other subjects in order to teach *BHH* (5) and they have much pressure to spend their time on reading, writing, and math (3). One teacher simply commented, "Takes time away from other things." A few teachers (12) said that taking part in *BHH* had not had an effect on their teaching of other subjects.

Uses of writing in teaching history

Teachers in treatment and comparison schools were asked to answer the question, "What kinds of writing about history have you asked your students to do?" Among the treatment teachers, 104 of the 115 teachers responded to this question for a response rate of 90%. Of those responding, about one-fifth (20 teachers), most of whom taught kindergarten or 1st grade, said that they did not ask their students to do any writing about history. The most common forms of writing that teachers said they asked students to do was writing short stories (33), writing letters (32), writing essays (including essay tests) (22), journals (17), drawing and sequencing (in younger grades) (10), photo analysis – labeling and descriptions (9), reflections on books or videos (5), and research reports and biographies (5). One teacher said that last year the students did journals, but this year they had done very little because they were not motivated to write "so doing the district's writing benchmarks was about all I could handle."

Among the 38 comparison teachers, 35 teachers responded for a response rate of 92%. Of the 35 responding, about one-third (11) of the comparison teachers said their students do very little or no writing as part of their history learning. Some of the writing activities that comparison teachers said that their students participated in were: essay or short answers (including essay tests) (15), short stories (7), letters (6), fact diagrams (4), drawing (3), learning logs (3), and journals (2).

Collaborative relationships

Treatment and comparison teachers were asked about collaborative relationships in teaching history. Treatment teachers were asked in what ways collaborative relationships that they had built as a result of the project were important to them. Of the 115 treatment teachers taking the survey, 102 responded to this question for a response rate of 89%. The most common ways that treatment teachers said that collaborative relationships were important were for sharing ideas and brainstorming (44), getting other perspectives on teaching the lessons (27), developing and reviewing lesson plans (20), and working together to gather and share materials (15). Some teachers (7) also said that they thought that working together on the units helped build a stronger team because they got to know each other better and a few (5) said that working together on history made teaching more fun. One teacher said collaboration was, "Extremely important. It's always great to glean ideas and what's worked or hasn't worked with particular activities embed in units" and another teacher said, "I think that this unit has brought the team of teachers together as a whole." Other comments were that they planned events together (3), collaboration helped create organizational consistency across the grade level (3), that collaboration helped them feel more prepared, have less concern that they were "missing something" (3) and several teachers (3) mentioned the value of collaborating with their *BHH* II mentor teacher.

Some teachers (17), particularly from one *BHH* II school, said that project participation had not changed the nature of collaboration with other teachers because they already had very good collaboration with their grade level peers. One teacher said, "Our team has always been a very cooperative one. This is a vital part of my job – it is support, encouragement, and friendship all rolled into one."

A few teachers (3) simply said that their collaboration has not changed as a result of the project and a few made other comments including, "We didn't do as much with collaboration this year – not time due to other requirements of district" and "I did the work as no one else was teaching the unit while I was."

The comparison teachers were asked in what ways they collaborated with other teachers in the teaching of history. Of the 38 comparison teachers surveyed, 32 teachers responded to this question for a response rate of 84%. More than half of the teachers who responded (17) said that they had collaborated very little or not at all with other teachers on teaching history. Some teachers said that they routinely collaborate with other teachers on many things including history (7), several teachers (5) gave specific details on their collaboration with other teachers such as, "1st grade teachers shared ideas of activities to do on Constitution Day", a few teachers (3) said they generally share ideas and materials, and a couple teachers (2) said they collaborated on history-related lesson plans, workshops, and pedagogy.

Perceived Benefits of the Curriculum Activities

The third part of the survey was administered to treatment teachers only and asked teachers to rate how beneficial each of the activities associated with the particular *BHH* curricular unit were for their students. The five-point scale used for these items ranged from "0" meaning "Not at all beneficial" to "5" meaning "Very beneficial." The item stems were identical to those used to examine thoroughness of implementation.

Kindergarten

History of Me

Treatment teachers responded to ten items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *The History of Me* were for their students. Nine of the ten of the activities were rated as at least a mean of 3.0 and six of the ten activities were rated 4.0 or above. The mean perceived benefit of all ten activities was 3.89.

Children Long Ago

Treatment teachers responded to six items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Children Long Ago* were for their students. All six of the activities were rated as at least a mean of 3.5 and three of the six activities were rated 4.0 or above. The mean perceived benefit of all six activities was 3.84.

First Grade

My First Grade History

Treatment teachers responded to seven items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *My First Grade History* were for their students. All seven of the activities were rated as at least a mean of 3.5 and six of the seven activities were rated 4.0 or above. The mean perceived benefit of all four activities was 4.37.

Community History

Treatment teachers responded to four items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Community History* were for their students. All four of the activities were rated as at least a mean of 3.5 and two of the four activities were rated 4.0 or above. The mean perceived benefit of all four activities was 4.16.

Second Grade

Immigration

Treatment teachers responded to eight items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Immigration* were for their students. All eight of the activities were rated as at least a mean of 3.0 and six of the eight activities were rated 4.0 or above. The mean perceived benefit of all nine activities was 4.21.

Environmental history

Treatment teachers responded to nine items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Environmental History* were for their students. All nine of the activities were rated as at least a mean of 3.5 and six of the nine activities were rated 4.0 or above. The mean perceived benefit of all nine activities was 4.08.

Third Grade

Industrialization

Treatment teachers responded to nine items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Industrialization* were for their students. Five of the nine activities were rated as at least a mean of 3.0, four of those were above 3.5. The four activities that teachers rated lower (range 2.33-2.93) were the same four that they said they had not implemented as thoroughly (see above).

Segregation

Treatment teachers responded to eight items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Segregation* were for their students. All eight of the activities were rated as at least a mean of 3.0 and six of the eight activities were rated 4.0 or above. The mean perceived benefit of all nine activities was 4.15.

Fourth Grade

The Progressive Era

Treatment teachers responded to eight items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on the *Progressive Era* were for their students. Seven of the eight activities were rated as at least a mean of 3.0 and six of the ten activities were rated 3.5 or above. The mean perceived benefit of all nine activities was 3.72. The activity that had a mean rating below 3.0, was "History of Capitalism" and received a mean rating of 2.5. Many of the economics concepts on which teachers said that few of their students

would be competent were associated with this activity. This activity was also rated as less thoroughly implemented by the teachers.

The Great Depression

Treatment teachers responded to nine items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on the *Great Depression* were for their students. All nine of the activities were rated as at least a mean of 3.0 and five of the nine activities were rated 3.5 or above. The mean perceived benefit of all nine activities was 3.72.

Fifth Grade

Native Americans

Treatment teachers responded to nine items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on *Native American History* were for their students. Five of the nine of the activities were rated as at least a mean of 3.0 and three of the nine activities were rated 3.5 or above. The mean perceived benefit of all nine activities was 3.27. There was greater variation between activities as to how beneficial they were in this unit than in many other units, with two activities, "Indian Removal and the Trail of Tears" and "Indian Boarding Schools: Policy and purposes" rated as quite beneficial (4.53 and 4.07, respectively) and two activities, "Assimilation Policy – The Dawes Act" and "Problems for contemporary Native Americans" rated somewhat low (2.67 and 2.73, respectively).

WWII – The Home Front

Treatment teachers responded to ten items on a five-point scale concerning how beneficial they believed different activities of the *BHH* curriculum on the *WWII Home Front* were for their students. All ten of the activities were rated as at least a mean of 3.5 and six of the ten activities were rated 4.0 or above. The mean perceived benefit of all nine activities was 4.3.

Summary of Perceived Benefit of the BHH Curriculum

Table 12 shows the treatment teachers' mean ratings across all items concerning how beneficial the *BHH* curriculum was for students. On the scale used, the anchor points were defined as "1" meaning "Not at all beneficial" and "5" as "Very beneficial", so scores indicate that the teachers thought that all units of the curriculum were beneficial and many thought it was very beneficial for their students to learn history with the *BHH* curriculum. Across all grade levels and topics, the treatment teachers' mean ratings ranged from 3.27-4.37. Six of the twelve units were rated between 3.27 and 4.0 and six mean ratings were greater than 4. The units

receiving the highest ratings on the benefit for students were *My First Grade History* (4.37), *WWII Home Front* (4.31), *Immigration* (4.21), *Community History* (4.16), *Segregation* (4.15), and *Environmental History* (4.08). Those rated the least beneficial (although still above a mean of 3 meaning the teachers felt it was beneficial) were the *Industrialization* (3.29) and *Native American History* (3.27) units.

In addition to the information concerning the benefits of the *BHH* II project gathered from the teacher implementation surveys, some treatment teachers were given the opportunity to talk about the benefits of the project during grade level focus group interviews conducted near the end of Year 3 of the project. [The focus group protocols and full transcripts are provided in the appendices to this report. The procedures for selecting and conducting the groups were reported in the Methods section of this report.] Teachers were asked to discuss the benefits for themselves as teachers, and their perception of the benefits to their students, of participating in the *BHHII* project. The following responses concerning benefits to the teachers emerged across several of the grade level groups:

- High quality curriculum and materials
- Excellent fit with existing curriculum
- *BHH* curriculum lends itself to making strong cross-curricular connections
- Teachers learned more history content
- Scope and content of the units were developmentally appropriate
- Curriculum helped relate history to their own and students' everyday lives
- Valuable opportunity to work with mentor teachers
- Teachers enjoyed teaching the curriculum and being part of the project

Other benefits mentioned by teachers included: helped teachers get to know students (and their families) better, timelines are good classroom tools, learned new pedagogical methods, and the curriculum stimulated important classroom discussions. One 1st grade teacher commented, "I guess the benefit is the scope and sequence of it all. How it just flows so nicely and usually it has vocabulary words that you are supposed to touch upon – it has books – I mean, it just has a variety of sources and resources to use to cover the content, so it's very teacher friendly."

The teachers thought the main benefits of the curriculum for their students were:

- Increased general awareness of history
- Increased interest in and motivation to learn history
- Helped students make cross-curricular connections
- Increased exposure to primary sources

- Encouraged students to read and learn more history on their own
- Increased empathy for people who lived in different times

One kindergarten teacher commented, "Maybe if our little folks get a little taste of history all the way through, then when they get into high school, they will have more connections for the history that's taught and perhaps they will like it better. Somewhere along the line, they stop liking history." A 3rd grade teacher said, "It was really an eye-opener for kids. I think it was hard for them to actually believe, especially in the segregation unit, that all those things happened. I think they were just amazed at some of the things they learned." In the 4th grade teacher focus group, one teacher said, "I saw students get excited about what they were learning and doing their own reading about the era because they were interested in it."

During the focus groups, treatment teachers were also asked what they thought were the negative consequences or least beneficial aspects for students of learning history with the *BHH* curriculum. Three of the six grade level focus groups did not have anything they thought was not beneficial for the students. Kindergarten teachers mentioned only that a couple of the activities were less interesting to the students – the music activity in the *History of Me* unit and the *Children Long Ago* activity on lights in history. They hypothesized that the lack of physical representation for these things in the form of artifacts made it more difficult for students to understand these things.

Some 4th grade teachers thought that the parts of the *Progressive Era* unit that were focused on economics more than history were more difficult and not as interesting to students as the rest of the unit. Fifth grade teachers thought the *Native American* unit was not age-appropriate for 5th grade students, required their classrooms to be overly "teacher-centered" and not as teacher-friendly as the *WWII Home Front* unit.

5.1.3 History Classrooms

Other history topics taught

Teachers were asked to list the history topics that they taught during the year (Treatment teachers were asked to list the topics they taught in addition to the *BHH* topics).

The full list of topics mentioned by *BHHII* treatment and comparison teachers is found in the Appendices to this report. A total of 86 of the treatment teachers replied to this item for a response rate of 75%. The topics most commonly mentioned by *BHHII* teachers were (in order of frequency of occurrence): Civil Rights, Presidents, Thanksgiving, Holidays, Revolutionary

War, Immigration, Native Americans, Explorers and Columbus, Colonies, and Communities (local history).

Among the comparison teachers, 30 of the 38 teachers responded to this item for a response rate of 79%. The most commonly listed topics by comparison teachers were: Slavery, Native Americans, Civil Rights, Presidents, Revolutionary War, Explorers and Columbus, Colonies, Thanksgiving, Changes in technology, Early Settlements, Immigration, Pioneers, and Regions of the US. There is considerable overlap between the two groups with eight of the most frequently listed topics mentioned by both groups. Most teachers at each grade level said they taught at least one other history topic. Of those who responded, eight treatment and two comparison teachers said they did not teach history topics.

Teachers were also asked to estimate the amount of time they spent on the topics associated with the *BHH* curriculum, and the amount of time they spent teaching other history topics. This was evidently a difficult task. In the treatment group, there was a somewhat reasonable range of responses for the amount of time that they spent teaching the *BHH* topics. For example, among treatment 5th grade teachers, their estimates of the time that they spent working on the *WWII Home Front* unit ranged from 40 minutes a day for 10 days to 40 minutes a day for 36 days. While this is a fairly wide range, teachers were allowed discretion to implement the curriculum as they preferred, and the average across classrooms (and probably the more typical) was about 40 minutes a day for 20 days.

Treatment teachers were even less consistent in their estimates of how much time they spent teaching *other* history topics, with the 5th grade teachers time estimates ranging from 40 minutes a day for 5 days to 40 minutes a day for 140 days. Teachers who teach other grades were equally varied in their estimates of time spent teaching history other than the *BHH* curriculum, ranging from 0 minutes, to 30 minutes every day.

Comparison teachers also varied widely in their time estimates on both tasks and sometimes there were contradictions in their responses. For example, one comparison teacher said that they spent 30 minutes a day, nearly every day of the year, studying the *BHH* topics but rated their thoroughness on the topics covered as a 1-2 in nearly all areas. In general, responses between treatment and comparison groups do not seem to differ greatly, with the exception that the amount of time treatment teachers said they spend on *BHH* topics appears to agree with their judgments of thoroughness of their instruction and is greater than the amount of time comparison teachers said they spent on the same topics. Teachers within the same school and grade level tended to respond in ways more like each other than like teachers from other schools, suggesting

that their grade level teams or school structure played at least some role in the amount of time teachers spent teaching history.

Observations

Summary of Observations of BHHII Classrooms

During the 2005-06 academic year, as part of the evaluation of the Bringing History Home II Project, evaluation team members observed *BHHII* participating classrooms that were in the process of using the *BHH* curriculum in their instruction. The observation visits were arranged ahead of time by the evaluators and teachers. Well in advance of the visits, teachers were asked to provide the evaluation team with a schedule of when they would be teaching the curriculum during a particular week and then evaluators chose the day to visit, based on convenience. The evaluator requested that the teachers use whatever lesson they had planned to teach that day and not rearrange their schedules to have special lessons on the days of the observations. The observations were intended to sample from the day-to-day routine of the curriculum in action, not just special or culminating days. Observations took place at all grade levels in all three participating *BHHII* school districts. [Classrooms in the longitudinal (*BHH* I) school were not observed because many visits were completed there during that grant period which were described in detail as part of the *BHH* I reports.]

The evaluation team members completed a total of 46 classroom visits; 29 in grades 3-5 and 17 in grades K-2. Table 14 shows the number of observation visits for each grade and at each school.

Table 14. Number of Classroom* Observations by Grade and School

Grade	Total Number of *BHHII* Classrooms	School 1	School 2	School 3	Total Number of Visits
K	17	2	4	0	6
1	17	0	3	2	5
2	16	2	0	4	6
3	16	4	2	0	6
4	15	0	6	3	9
5	15	5	5	4	14
All Grades	96	13	20	13	46

*In all but four cases, the number of classrooms is the same as the number of teachers. Four classes were team-taught by two teachers.

Several factors affected the number of classrooms visited at each grade and school: convenience (one school was substantially closer than the other two), weather (one visit had to be

cancelled because of inclement weather and it was not possible to reschedule), scheduling difficulties (some grade levels did not coordinate their teaching with other teachers, some taught units in very small time increments (15 minutes), and at some schools all the grade level classes received instruction on the same topic at the same time), cost (it was not cost effective to make observation visits for one class period, especially if the instructional time was very brief), and difficulty in communication with grade level contacts at some schools (one teacher from each grade level at each school was designated to work with evaluators on scheduling visits, but some were more helpful than others). The primary intent of the observations was not to monitor or describe differences between schools or teachers, but to describe and discuss what the curriculum looked like in a wide range of classrooms across the project.

In addition to being distributed across three schools and six grades, the *BHHII* Project involves 12 different sets of curriculum. Each grade level has two sets of curricula, intended typically to be taught in the fall and spring of the year. Observations were conducted primarily in classrooms during the fall of 2005, so some sets of curricula were not observed. Table 15 shows the distribution of observations over the different *BHH* units.

Table 15. Number of Classroom Observations by Grade and Curriculum Unit

Grade level/Curriculum unit	Number of observations
K/History of Me	0
K/Children Long Ago	6
1/My 1st Grade History	5
1/Communities Long Ago	0
2/Immigration	6
2/Environmental History	0
3/Industrialization	4
3/Segregation	2
4/Progressive Era	2
4/Great Depression	7
5/Native American	14
5/WWII Home Front	0
TOTAL	**46**

The same procedure was used for all observations. Evaluation team members arrived shortly before the teacher began teaching the *BHH* curriculum, spoke briefly with the teacher, asking where it would be best to sit to observe, and then acted as primarily passive observers, taking notes and completing an observation protocol (attached as Appendices to this report). On several occasions, students were engaged in independent or group activities so that in order to observe what was going on it was necessary for the evaluator to move about the room and occasionally during that process, students asked questions or had brief opportunities to talk with evaluators.

Three evaluators observed classrooms. The lead evaluator did 35 of the observations and the other two evaluators did seven and four observations. The lead evaluator worked with the two other evaluators to standardize the observation scores by observing at least one classroom together and then by post-observation discussion of the classes the others had visited to determine how to code activities and student and teacher engagement. Evaluators began the observation protocols during the visit and completed the protocols as soon as possible after each observation.

In order to systematically describe the classroom activities, an observation protocol was used that required observers to classify the types of activities that took place during the class, quantify the amount of time spent on each activity, make judgments as to the students' and teachers' level of engagement during the activities, and write a brief description of the main classroom activities.

The main purpose of the observations was to determine the extent to which students in *BHHII* classrooms were actively engaged in learning history. When all observations were complete, the evaluation team established a holistic rubric for assigning a score to each observation. The rubric employed a five-point scale based on several dimensions. The dimensions are:

- Student activity – extent to which students are actively learning history
- Student engagement – extent to which students appear engaged in learning history
- Stimulation for historical thought – extent to which the classroom situation (curriculum, materials, and teacher) support student use of historical thinking skills
- Teacher engagement – extent to which teachers display enthusiasm and interest in teaching history

An additional dimension was observed; the extent to which the lesson observed was a history lesson and/or used the *BHH* curriculum. There was little variation among classrooms on this dimension because nearly all classrooms used the curriculum (or history-based teacher modifications of the lesson) for the entire class period, therefore this dimension was not considered in the final holistic scoring of observations.

For scoring, the observation protocols were divided into two groups; kindergarten through 2nd grade, and 3rd grade through 5th grade. There were two reasons for grouping the observation data in this way. Using the larger groups provides more bases for looking across classrooms than using the limited number at each grade, and the curricular formats and activities used in the 3rd through 5th grade are more similar to each other than they are to the younger grades curricula. Although the 2nd grade curriculum has elements of both the younger and older grade's

curricular formats, the nature of the activities carried out in 2nd grade is more like the younger grades, so the 2nd grade observations were grouped with kindergarten and 1st grade observations.

The lead evaluator read all the observation protocols and used the dimensions of the rubric to sort observation protocols into five levels, 1-5, where a score of 5 means:

- Students are *actively* learning history
- Students are *highly engaged* in learning history
- Students are provided with *multiple opportunities* to use and exhibit use of historical thinking skills
- Teachers display *high enthusiasm and interest* in history instruction

A score of 1 means:

- Students are *passive* or are not involved in learning history
- Students are *not engaged* in learning
- Students are given *few or no opportunities* to use and exhibit historical thinking skills
- Teachers *do not exhibit enthusiasm or interest* in history instruction

Scores of two to four mean that the ratings of at least some of the dimensions fall somewhere in-between those two extremes. In the classrooms assigned mid-range holistic scores, it is possible that one or more dimension could be rated very high or low, but the overall impression would preclude the classroom observation from earning the extreme scores.

After reading all protocols (including descriptions of the day's activities as well as observer summaries and scores), assigning preliminary holistic scores, and then re-reading and making any adjustments as necessary, the classroom observations were categorized into five groups based on the holistic score. These scores are shown in Table 16.

Table 16. Distribution of Observation Protocol Scores on 5-point Rubric by Grade Level Group

		Observation Protocol Rubric Scores				
Grade level group		**1**	**2**	**3**	**4**	**5**
Grades K-2	**N**	2	1	4	6	4
	(%)	*(12)*	*(6)*	*(24)*	*(36)*	*(24)*
Grades 3-5	**N**	2	1	6	12	8
	(%)	*(7)*	*(3)*	*(21)*	*(41)*	*(28)*

After the protocols had been grouped as described in Table 16, each observation protocol was also looked at analytically using the following values that were determined by the evaluator/observer at the time of (or soon after) the observation:

1. Student Activity: Proportion of time that student activity is embedded in instruction (0-100%)

2. Historical thinking skills: Proportion of time that students are using historical thinking skills (see operational definition below) (0-100%)
3. Student engagement: Summary of proportion of students who appear engaged in learning during class period (coded on a 6-point scale ranging from "0" for "None" to "5" for "All" of the students)
4. Teacher engagement: Summary of general level of apparent teacher engagement in teaching history (coded on a 5-point scale from "0" for "Not at all" to "4" for "High")
5. Active Environment for Learning History: Overall estimate of the proportion of students who are actively learning history during the class period (coded on a 5-point scale from "0" for "None" to "4" for "All of the classroom")

The operational definition for historical thinking skills that was used for completing the observation protocols was derived from the goals outlined by the Bringing History Home Project. The definition included, but was not limited to the following observable behaviors: analyzing primary sources (such as photographs or written documents), creating or using timelines, creating or using historical narratives, using historical content with maps, creating mind maps, comparing different interpretations of history, comparing and contrasting to define historical concepts, and historical simulations.

Table 18 and 19 shows how the analytical scores assigned at the time of the observation aligned with the holistic scores assigned after review of all the protocols.

Since the observations took place in a wide variety of settings (three schools, six grade levels, and at different points of time during the implementation of eight different curriculum units), it is difficult to talk in generalities about the curriculum implementation. Table 17 shows the distribution of the observation protocol scores across the *BHH* different curriculum units.

Table 17. Distribution of Holistic Rubric Scores by BHH Curricular Unit

	Holistic Rubric Score				
Curriculum Unit	**1**	**2**	**3**	**4**	**5**
Children Long Ago(K)	0	0	3	2	1
My 1st Grade History (1)	1	1	1	0	2
Immigration (2)	1	0	1	3	1
Industrialization (3)	0	0	1	2	1
Segregation (3)	0	0	0	1	1
Progressive Era (4)	0	0	0	2	0
Great Depression (4)	0	0	2	2	3
Native American (5)	2	1	3	5	3

Table 17 shows that, for most of the curricula observed, the experiences of students in the *BHHII* classrooms varied, but most of the classrooms received a rating of three or higher on the

holistic rubric and at least half of the classrooms at each grade level were rated as four or five. For two sets of curricula, the 5th grade *Native American* unit and the 1st grade *My 1st Grade History* unit, scores varied across the entire range, from one to five.

Within the *BHH* curriculum, there is variation in the intrinsic student activity level. For example, some days call for the teacher to read aloud to students, a valued practice in a classroom, but one that does not always appear vibrant and one that may vary greatly as teachers' delivery of the materials may involve more or less opportunity for active discussion and even the content of the discussion may be more or less interesting to students. Because classroom visits were effectively randomly sampled, some visits may have occurred on the days that were least inherently active and others may have occurred on the liveliest. The differing experiences of 5th grade classrooms using the Native American curriculum will be examined further in future case study analysis.

Table 18. Comparison of Holistic and Analytic Rubric Scores for 3rd -5th Grade Observation Protocols

		Analytic Component Scores				
Holistic Rubric Scores	**# assigned each score**	**Student activity**	**Historical thinking skills**	**Student engagement**	**Teacher engagement**	**Active environment for learning history**
5	**8**	All observations rated students as active 90-100% of the time.	Nearly all of the observations said that historical thinking skills were used 90-100% of the time.	All observations were rated as nearly all or all of the students were engaged in learning (4.5-5).	All observations said that teachers displayed enthusiasm and interest in instruction all or nearly all of the time (4)	The overall summary of the extent to which the classrooms were engaged in actively learning history ranged from "much" to "all" (3-4).
4	**12**	Most observations said that students were active 90-100% of the time, but several were scored as 80%.	Observations ranged from 20-100%, with a median score of 90%.	Observations were rated as most or all of the students were engaged in learning (4-5).	Observations said that teachers displayed enthusiasm and interest in instruction much to all of the time (3-4).	For all of these observations, the overall summary was that "much" of the classroom was actively engaged in learning history (3).
3	**6**	The range of student activity ranged from 40-100%, with a mean of 60%.	Observations ranged from 40-100% with a median score of 65%.	Observations were all rated that most students were engaged (4).	Observations said that teachers displayed enthusiasm and interest some to all of the time (2.5-4.0) with a median rating of 3.	The overall summary of the extent to which the classrooms were engaged in actively learning history ranged from "some" to "much" (2-3).
2	**1**	(Only one observation was rated a "2") In this category it was scored as 100%.	In this category it was scored as 60%.	In this category it was scored as "most" (4).	In this category it was scored as "some" (3).	In this category it was scored as "some/much" (2.5).
1	**2**	Both observations received scores of 10% for this category.	One observation received a score of 0% and the other 10%.	In one of these classrooms, student engagement was rated as "4" the other as "3".	One teacher was rated as displaying enthusiasm and interest some of the time and the other as little of the time (1-3).	The overall summary of the extent to which the classrooms were engaged in actively learning history ranged "little" to "some" of the time (1-2).

Table 19. Comparison of Holistic and Analytic Rubric Scores for Kindergarten – 2nd Grade Observation Protocols

		Analytic Component Scores				
Holistic Rubric Scores	**# assigned each score**	**Student activity**	**Historical thinking skills**	**Student engagement**	**Teacher engagement**	**Active environment for learning history**
5	**4**	Most of the observations rated students as active 100% of the time.	Most observations said historical thinking skills were used during 90-100% of the time.	All observations were rated nearly all or all students were engaged. (5)	Most of the teachers were rated as showing high enthusiasm in teaching. (4)	The overall summaries of the extent to which the classrooms were engaged in actively learning history ranged from "much" to "all/much" (3-3.5).
4	**6**	Most of the observations rated students as active 100% of the time.	Ratings of historical thinking skills used ranged from 50-100% with a median of 87%	Half of these observations rated most students as being engaged and the other half said all students were engaged. (4-5)	Most of the teachers were rated as displaying much enthusiasm and the rest were rated as displaying high enthusiasm. (3-4)	The overall summary of the extent to which the classrooms were engaged in actively learning history was "much" for all observations. (3)
3	**4**	Ratings for student activity ranged from 50-100% with a median of 82%.	Ratings of historical thinking skills used ranged from 50-100% with a median of 87%.	The ratings in student engagement ranged from "about ½" to all (3-5) with a median of 4.25.	The rating of teacher enthusiasm ranged some to much (2.5 to 3.5) with a median of 3.	The overall summaries of the extent to which the classrooms were engaged in actively learning history ranged from "some" to "much" (2-3) with a median of 2.75.
2	**1**	Students were rated as active 90% of the time.	The rating for use of historical thinking skills was 75%.	The students' level of engagement was rated as "most" (4).	The teacher was rated as exhibiting "some" to "much" enthusiasm (2.5)	Overall, this class was seen as involving student activity in learning history "some" to "much" of the time. (2.5)
1	**2**	Students were rated as active 100% of the time.	Ratings of historical thinking skills used ranged from 40-50%.	The students level of engagement was rated as "about ½" to "much" (3-4).	The teachers were rated as exhibiting "some" to "much" enthusiasm (2.5)	Overall, these classes were seen as involving student activity in learning history "some" of the time (2).

5.2 Student Outcomes

5.2.1 Teacher Perception of Student Competencies

The second part of the teacher implementation survey asked teachers to estimate the proportion of their students who they believed were competent at performing knowledge and skill tasks associated with each of the *BHH* curriculum units. The scale used for these items ranged from "0" meaning "None are able to do this competently" to "5" meaning "All or nearly all can do this competently." In most cases, the item stem reflected the title of content knowledge or historical thinking goals as specified in the *BHH* curriculum.

Kindergarten

History of Me

Teachers responded to 19 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *History of Me* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 15 of the 19 tasks (a mean score of 3.0 or above). The mean across all 19 items for treatment teachers was 3.53. On each of the tasks, the treatment teachers' group mean was greater than the comparison teachers' group mean and on some of the tasks, the score distribution was nearly non-overlapping. There was one task ("Tell something about their own histories using pictures or artifacts") for which the majority of treatment teachers thought that all or nearly all of their students could perform the task competently. Other tasks that treatment teachers said many of their students could perform competently were, "Describe how human bodies change over time (growth)" (treatment mean =4.27), "Tell where they were born" (treatment mean=3.91), and "Put pictures of themselves and other artifacts that they had when they were babies, toddlers, and kindergartners in sequence" (treatment mean =4.20).

The mean rating on 17 of the 19 tasks for the comparison group teachers was 3.0 or less and for 7 of the 19 tasks the mean was less than 2.0. The mean across all tasks for comparison teachers was 2.29.

Children Long Ago

Teachers responded to 12 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Children Long Ago* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 11 of the 12 tasks (a mean score of 3.0 or above). The mean across all 12 items for treatment teachers was 3.66. On each of the 12 tasks, the

treatment teacher group mean was greater than the comparison teacher group mean and on most tasks the score distribution was nearly non-overlapping. The tasks where there were the largest differences between treatment and comparison teachers' perception of their students' competence were those that called upon students to compare and contrast different aspects of life today and life long ago.

The mean rating on all of the 12 tasks for the comparison group teachers was 3.0 or less and for 11 of the 12 tasks the mean was less than 2.0. The mean across all tasks for comparison teachers was 1.78. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

First Grade

My First Grade History

Teachers responded to 18 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *My First Grade History* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing all 18 tasks (a mean score of 3.0 or above) and they said that two-thirds of their students could perform 8 of the18 tasks competently. The mean across all 18 items for treatment teachers was 3.83. On three of the tasks, the majority of the teachers said that "All or nearly all" of their students would be competent at performing the task. These tasks included, "Tell stories about their own lives", "Describe the difference between a true story and a make-believe story" and "Put school day or week events in sequence". Other tasks on which the treatment teachers rated their students' competence quite high included:
"Describe the meaning of the word 'history'" (treatment mean=4.20), "Tell a story using pictures, letters, or artifacts" (treatment mean=4.24), "Create a timeline of their day or week at school" (treatment mean=4.00), "Use photos or art to represent activities" (treatment mean=4.36), and "Describe how maps are useful to people" (treatment mean=4.24). On each of the 18 tasks, the treatment teacher group mean was greater than the comparison teacher group mean.

The mean rating on 13 of the 18 tasks for the comparison group teachers was 3.0 or less and on 7 of the 18 tasks, the mean rating was less than 2.0. The mean across all tasks for comparison teachers was 2.61.

Community History

Teachers responded to ten items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Community History* unit. The treatment teachers' responses indicated that they believed at least one-third of

their students were competent at performing all ten tasks (a mean score of 3.0 or above). The mean across all ten items for treatment teachers was 3.87. On three of the tasks, the majority of the teachers said that "All or nearly all" of their students would be competent at performing the task. These tasks included, "Explain or draw a picture of what their community looked like long, long ago", "Describe some of the differences in their community 'long, long ago', 'long ago', and 'today'" and "List some of the services that towns usually have (e.g., police, schools, library, post office, stores". On each of the ten tasks, the treatment teacher group mean was greater than the comparison teacher group mean.

The mean rating on eight of the ten tasks for the comparison group teachers was 3.0 or less. The mean across all tasks for comparison teachers was 2.45. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

Second Grade

Immigration

Teachers responded to 17 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Immigration* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 15 of the 17 tasks (a mean score of 3.0 or above) and six of the 17 tasks received a mean rating of 4.0 or above. The mean across all 17 items for treatment teachers was 3.67. On each of the 17 tasks, the treatment teacher group mean was greater than the comparison teacher group mean, sometimes with non-overlapping or nearly non-overlapping score distributions. On two of the tasks, "Define the word immigrant" (treatment mean=4.50) and "Describe Ellis Island and tell what it was like for immigrants to come there" (treatment mean=4.69), at least half of the treatment teachers said, "All or nearly all of their students can perform this task competently". Other tasks on which treatment teachers rated their students' competence high were: "Explain how many of our ancestors were immigrants to the US or were nomadic people who came to the US from other locations" (treatment mean=4.19) and "Describe some of the reasons that people leave their country of birth. Cite examples from early 20th century US immigration" (treatment mean=4.19).

The mean rating on all 17 tasks for the comparison group teacher was 3.0 or less, and on 14 of the 17 tasks the mean was 2.0 or less, meaning teachers believed only a few of their students would be able to perform the specified tasks competently. The mean across all tasks for

comparison teachers was 1.57. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

Environmental history

Teachers responded to 26 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Environmental History* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 23 of the 26 tasks (a mean score of 3.0 or above). The mean across all 26 items for treatment teachers was 3.53. On each of the 26 tasks, the treatment teacher group mean was greater than the comparison teacher group mean. Examples of tasks on which the differences between treatment and comparison teachers' ratings of student competency were greatest were (with treatment mean followed by comparison mean in parentheses):

- Discuss the work of some early environmentalists (Roosevelt, Muir, or Pinchot) (3.45, 1.00)
- Create and place appropriate depictions of logging, farming, and mining on a map of the US (3.48, 1.17)
- Describe the harmful environmental byproducts of manufacturing processes down through the years (3.31, 1.14)
- Discuss Rachel Carson and the purpose and message of her book *Silent Spring* (3.20, 1.00)

The mean rating on 21 of the 26 tasks for the comparison group teacher was 3.0 or less, and on 16 of the 26 tasks was 2.0 or less meaning they believed only a few of their students would be able to perform the specified tasks competently. The mean across all tasks for comparison teachers was 1.96. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

Third Grade

Industrialization

Teachers responded to 21 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Industrialization* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 13 of the 21 tasks (a mean score of 3.0 or above). The mean across all 21 items for treatment teachers was 3.22. On 19 of the 21 tasks, the treatment teacher group mean was greater than the comparison teacher group mean. On the two tasks on which the comparison mean exceeded the treatment mean, the ratings of student

competencies were somewhat low in both cases and the differences were small. The two items were: "Describe some of the reasons for recent immigrations to the US" (comparison mean=2.67, treatment mean=2.43) and "Define the term 'investors'" (comparison mean=2.17, treatment mean=2.00).

The mean rating on 17 of the 21 tasks for the comparison group teacher was 3.0 or less, and on 11 of the 21 tasks was 2.0 or less meaning they believed only a few of their students would be able to perform the specified tasks competently. The mean across all items for comparison teachers was 2.2. It is important to note that in the scale used on these items, a score of "1" means that the teacher thought that none of their students could perform the task competently.

Segregation

Teachers responded to 19 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Segregation* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 14 of the 19 tasks (a mean score of 3.0 or above). The mean across all 19 items for treatment teachers was 3.3. On each of the 19 tasks, the treatment teacher group mean was greater than the comparison teacher group mean.

Examples of tasks on which the differences between treatment and comparison teachers' ratings of student competency were greatest were (with treatment mean followed by comparison mean in parentheses):

- List some of the Jim Crow laws (3.56, 1.00)
- Describe what life was like for African Americans during the time of the Jim Crow laws (3.78, 1.17)
- Explain what the 13th Amendment is about (3.78, 1.83)
- Describe some ways that African American people resisted Jim Crow laws (3.17, 1.00)
- Describe the main purpose of the Emancipation Proclamation (3.22, 1.50)

The mean rating on 16 of the 19 tasks for the comparison group teacher was 3.0 or less, and on 11 of the 19 tasks was 2.0 or less meaning they believed only a few of their students would be able to perform the specified tasks competently. The three tasks that comparison teachers believed one third or more of their students could perform competently were, "Describe some different kinds of rules in the US" (3.33), "Define the word prejudice" (3.08), and "Define the word segregation" (3.08). The mean across all tasks for comparison teachers was 1.96. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

Fourth Grade

The Progressive Era

Teachers responded to 23 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Progressive Era* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 14 of the 23 tasks (a mean score of 3.0 or above). The mean across all 23 items for treatment teachers was 3.3. On each of the 23 items, the treatment group mean was greater than the comparison group mean. There was a fair amount of variation between items on the teachers' rating of student competencies with competence at some activities associated with immigration, child labor, and urban poor rated quite high, such as "Define immigration" (4.14), "Identify some of the reasons that many immigrants worked in factories" (4.07), "Identify some of the reasons that children worked in factories" (4.29), "Describe what it was like for a child to work in a factory" (4.43) "Analyze historic photographs for information" (4.25) and "Describe what housing was like for poor people in US cities at the end of the 19th and beginning of the 20th centuries" (4.21). On other items, having primarily to do with economics, such as "Describe the roles of profit and competition in a capitalist system from a historical perspective" (2.23), "Define the term 'corporation'" (2.62), "Define the term 'capitalism'" (2.69), and modern progressive action items such as "Define the term 'regulatory law' and give an example" (1.83) and "Describe a progressive project that could be or is being done today" (2.45), teachers rated student competency quite low.

The mean rating on 20 of the 23 tasks for the comparison group teacher was 3.0 or less, and of 14 of the 23 tasks the mean was less than 2.0, meaning they believed only a few of their students would be able to perform the specified tasks competently. The mean rating of student competency across all tasks for comparison teachers was 1.94. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

The Great Depression

Teachers responded to 18 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Great Depression* unit. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 13 of the 18 tasks (a mean score of 3.0 or above). The mean across all 18 items for treatment teachers was 3.3. On each of the 18 tasks, the treatment group mean was greater than the comparison group mean.

Some of the topics for which the treatment teachers' perception of student competencies differed most from comparison teachers' perception of their students' competencies were (with treatment mean followed by comparison mean in parentheses):

- Describe what life was like for a child during the Depression (4.12, 1.75)
- Compare and contrast life in the 1920's and 1930's (3.82, 1.38)
- Describe typical clothing and housing during the Depression (3.88, 1.50)
- Analyze a historical photograph for information (3.82, 2.00)
- Describe a "penny auction" and explain its purpose (3.76, 1.00)
- Describe unemployment during the Depression (3.67, 1.38)

The mean rating on 14 of the 18 tasks for the comparison group teacher was 2.0 or less, meaning they believed only a few of their students would be able to perform the specified task competently. The mean across all tasks for comparison teachers was 1.7. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

Fifth Grade

Native Americans

Teachers responded to 20 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *Native American History* unit. There were only five comparison teachers surveyed, so comparisons between the two groups are necessarily limited, but the magnitude of the differences between the two groups on most items in this section appears meaningful. The treatment teachers' responses indicated that they believed at least one-third of their students were competent at performing 12 of the 20 tasks (a mean score of 3.0 or above). The mean across all 20 items for treatment teachers was 3.2. On each of the 20 tasks, the treatment group mean was greater than the comparison group mean.

Examples of tasks on which the differences between treatment and comparison teachers' ratings of student competency were greatest were (with treatment mean followed by comparison mean in parentheses):

- Describe the "Trail of Tears" (4.33, 1.40)
- Take a position either supporting or opposing the Removal Policy (4.00, 1.60)
- Compare and contrast Native American's live on and off of the Indian Reservations (3.40, 1.20)

- Write a narrative of what it might have been like to be a Native American boarding school student (3.67, 1.40)
- Define the term "Indian Reservation" (3.93, 2.00)

The mean rating on 19 of the 20 tasks for the comparison group teacher was less than 3.0, and on 16 of the 20 tasks the mean was below 2.0, meaning they believed at most a few of their students would be able to perform the specified tasks competently. The mean across all tasks for comparison teachers was 1.65. Again, it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

WWII – The Home Front

Teachers responded to 17 items on a 5-point scale concerning their perception of their students' competencies at performing content knowledge or skills tasks related to the *WWII Home Front* unit. There were only four comparison teachers surveyed, so comparisons between the two groups are necessarily limited, but the magnitude of the differences between the two groups on most items in this section appears meaningful. The treatment teachers' responses indicated that they thought at least one-third of their students were competent at performing 13 of the 17 tasks (a mean score of 3.0 or above) and at least two-thirds of their students were competent at performing nine of the 17 tasks (a mean score of 4.0 or above). The mean across all 17 items for treatment teachers was 3.7. On each of the 17 tasks, the treatment group mean was greater than the comparison group mean and often the score distribution was nearly non-overlapping.

There were several tasks that the majority of the treatment teachers thought "All or nearly all" of their students would be able to perform competently. These tasks were:

- Define the term "home front"
- Define the term "war front"
- Describe how the role of women in the workplace changed as a result of WWII
- Describe some ways that citizens on the US Home Front helped the war effort during WWII
- Describe how Japanese Americans were treated in the US during WWII

Other tasks on which the teachers rated their students' competencies high included: "Describe how the need for factories changed as a result of WWII" (treatment mean=4.1), "Describe what it was like for families to have someone serving in the military during WWII" (treatment mean=4.0), "Explain how rationing was used during WWII" (treatment mean=4.3), "Describe what life was like for Japanese in internment camps" (treatment mean=4.1) and "Create a mind map about the WWII Home Front" (treatment mean=4.0).

The mean rating on all 17 tasks for the comparison group teacher was less than 2.0, meaning they believed at most a few of their students would be able to perform the specified task

competently. The mean across all tasks for comparison teachers was 1.6. Again it is important to note that "1" on this scale means that the teachers thought that none of their students would be able to perform the task competently.

Summary of Teacher Perception of Student Competencies

Table 12 shows the teachers' mean ratings of student competencies at performing kills or displaying knowledge associated with the *BHH* curriculum. Across all grade levels and topics, the mean rating of student competencies given by treatment teachers was greater than the comparison teacher mean. The mean ratings by treatment teachers ranged from 3.22 to 3.8, indicating that treatment teachers believed between one-third and two-thirds of their students were competent at performing the knowledge and skills tasks indicated.

For comparison teachers, the range of mean ratings of students' competencies was 1.57 to 2.61, with only four topics having a mean rating greater than 2.0 (Kindergarten *History of Me* (2.29), *My 1st Grade History*(2.61), 1st grade *Community History* (2.45) and 3rd grade *Industrialization* (2.20). Therefore, in general, mean ratings indicated that comparison teachers thought that at most only a few of their students could perform the tasks specified. The largest differences between the treatment and comparison teachers in their perception of student competencies occurred in the *Immigration* (2nd grade), *WWII Home Front* (5th grade), and *Children Long Ago* (Kindergarten) units.

In addition to the scaled items concerning teachers' perceptions of their students' ability to perform certain tasks or display specific knowledge regarding the topics studied, teachers were also asked several open-ended questions concerning their perception of what students know or can do and what they think they *should* know or be able to do. First teachers were asked, "Other than those listed above (in the scaled items), what kinds of historical knowledge or skills did your students learn during this school year that you believe they would be at least 'somewhat competent' at performing?" When looking at teachers' responses to this question, the teachers' responses were divided into the two main categories of "Knowledge" and "Skills". When categorized this way, a definite pattern emerged. The knowledge that the teachers thought that their students' would be able to display competently was similar across the two groups with the top ten topics having quite a bit of amount of overlap, especially in the topics most frequently mentioned. The most popular topics listed by treatment teachers were (with comparison teacher rank in parentheses):

1. African-American history (slavery, segregation, civil rights, Martin Luther King, Jr., Harriet Tubman) (1)
2. Revolutionary War and independence from England (2 tied)

3. (tied) Native Americans (2 tied)
3. (tied) Presidents: Lincoln, Washington (4 tied)
3. (tied) Colonization, pilgrims (7)
4. (tied) Immigration (8)
4. (tied) National Holidays (9)
4. (tied) Personal history (the child's or family history)
4. (tied) Westward Movement – pioneers, settlers (4 tied)

The other topics listed most often by comparison teacher were (with ranking in parentheses): Civil War (6), Explorers (8 tied), and Citizenship (8 tied). The complete list of topics listed by treatment and comparison teachers is found in the appendices to this report. Where treatment teachers and comparison teachers responded differently was in the presence or absence of listing skills that their students would be competent at performing. Treatment teachers listed many different skills that they thought their students possessed, where only a few comparison teachers listed skills at all and they were limited to only three skills. The skills most commonly listed by treatment teachers were (with the number of teachers mentioning it in parentheses):

- Compare and contrast things from long ago and things from today (sometimes general, sometimes given with specifics e.g., schools, houses, children's lives) (7)
- Timelines (identify, create, explain that it shows the passage of time) (7)
- Map usage (analyze, create, locate things) (5)
- Photo analysis (also, matching old and current photos of the same thing) (4)
- Understand more about other times and cultures (3)
- Examine cause and effect in history (sometimes general, sometimes given with specifics e.g., reasons for American Revolution and Civil Wars and impact, affect of Martin Luther King, Jr.'s life on African-Americans)(2)
- Relate to the past (2)
- Awareness of what we can learn from history's successes and failures (2)

In addition to those skills, single treatment teachers mentioned other skills including, generating questions, research skills, writing narratives, understanding continuity and change over time, and drawing conclusions. The only skills mentioned by comparison teachers were timeline usage (3), map or globe usage (3), and examining cause and effect (1). It appears that when teachers were asked the double-barreled question concerning student competence at knowledge and skills, comparison teachers thought primarily of the content knowledge that they had taught to their students, while treatment teachers thought of both content knowledge and the skills that students acquired to help them learn and understand history.

Teachers were also asked to identify the most *important* history-related knowledge or skills that their students should know or be able to perform. Again treatment teachers were much more likely than their comparison group peers to list *skills* related to learning history. The skills listed most commonly by treatment teachers are similar to the ones at which they said that their students were becoming competent (with the frequency of the response in parentheses):

- Use of timelines – understanding chronology and sequencing (19)
- Primary source analysis – documents, photos, artifacts (19)
- Comparing and contrasting – old vs. new, now vs. long ago (15)
- Drawing conclusions, making inferences (4)
- Understanding connections between past and present (4)
- Understanding cause and effect (3)
- Making mind maps to organize knowledge (3)
- Generating good questions (3)
- Doing research (3)
- Empathy for people in historical times (2)

In contrast, only a few of these were mentioned by comparison teachers (with frequency of responses in parentheses), understanding cause and effect (3), use of timelines (2), and using map skills (2), and several of these were all mentioned by one teacher. Again, the responses suggest that comparison teachers are more apt to think of history learning in terms of content knowledge over skill acquisition.

In addition, the types of topics that teachers mentioned as being important for elementary students to know about were somewhat different in the treatment and comparison group with treatment teachers more likely than comparison teachers to mention general topics such as, understanding how things change over time, understanding what "history" is, general knowledge of the world, and understanding that you can share your own history. They also mentioned general topics such as local history, colonization, the Civil War, and Segregation among their most important topics. Comparison teachers' most common choices of topics included Founding Fathers (famous Americans), Explorers and Columbus, the Civil War, Native Americans, and general knowledge of US History.

During focus group interviews with grade level treatment teacher groups, treatment teachers also talked about some of the knowledge and skills that they thought their students had gained as a result of the *BHH* curriculum. These included (in no particular order):

- Better understanding of what "history" is

- Increased content knowledge
- Ability to think chronologically
- Developed empathy
- Increased vocabulary
- Enhanced ability to present knowledge to classmates
- Learned to look for solutions
- Learned to generate good questions
- Developed research skills

One teacher said, "I think it increased their questioning and research skills. They wanted to find out more information." Another commented, "The kids were very empathetic. They…*we* had lots of fun. The kids were really pretty fascinated with a lot of the aspects of the different things and I think they learned a lot."

5.2.2 Narrative Assessments

Written narrative assessments concerning the six *BHH* content areas implemented in 3rd through 5th grades were collected from students as one of three types of written assessments intended as direct means of examining student learning outcomes that occurred as a result of the *BHH II* project. Assessments asked students to construct narratives using seven key words (for 3rd Grade Industrialization there were six words) from the content area they were studying. Resulting narratives were scored with 0, 1, or 2, points possible for each key word used, so the maximum score possible on all narrative assessments was 14 (12 for the 3rd Grade Industrialization assessment). Estimates of reliability of these scores in the form of Cronbach's alpha for each grade level and content area are shown in Table 5 in the Methods section of this report.

Treatment and comparison classrooms in which students completed the narrative assessment were assigned to either a *pre-post* or *post only* condition. Approximately half of the classrooms completed both a pre and a posttest (identical test for pre and post administration) and the other half completed the posttest only. Scores on all posttests were pooled for this analysis. Differences between [The effects of the testing condition (pre-post vs. post only) were minor are not included in this report.] The effects of treatment condition on student performance on the narrative measures are summarized in Table 20. Data are reported at the student level. Future research will include analyses of classroom level data.

Table 20 presents the important comparisons between treatment students on narrative pre and posttests and between treatment and comparison students. The table shows means for pre and

posttest performance (out of a possible 14 points, 12 points for the 3rd Grade Industrialization narrative assessment) along with standard deviations, standard errors, range of scores, and 99% confidence intervals for each test and treatment condition. Comparison of group means between two groups with non-overlapping intervals indicate 99% likelihood that the difference between the two means is significant.

As can be seen in Table 20, students in treatment schools on average scored higher on narrative posttests than they did on pretests indicating that growth in historical content knowledge occurred over the course of the project. In addition, students in treatment schools scored higher on the narrative posttests than did comparison school students over the same school year, indicating that the content knowledge was not something that children of the same grade typically learned during the school year. Treatment students and comparison students did not differ on narrative pretest scores and in some cases it appears that comparison students' pretest scores were slightly higher (although not significantly) than pretest scores of treatment students. Increases in treatment students' scores from pretest to posttest occurred at all grade levels and for all units, with increases in group means ranging from 4.3 to 5.8 points. In contrast, among comparison students, the largest gain in group means between pre and posttests was less than 1 point. All differences between treatment students' mean pre and posttest scores were significant as illustrated by non-overlapping 99% confidence intervals and (using the same criteria) all differences between treatment and comparison students' mean posttest scores were also statistically significant.

Table 20. Comparisons of Pre/Post and Treatment/Comparison Conditions for Student Narrative Assessments

Grade/Topic Test Condition (n_s/n_c)	Mean (k=7)*	SD	SE	Range	99% Confidence Interval
3rd Grade/Industrialization (k=6)*					
Pretest					
Treatment (45/3)	1.67	1.78	0.27	0-8	0.95-2.38
Comparison (24/2)	3.21	2.47	0.50	0-8	1.79-4.62
Posttest					
Treatment (65/5)	7.37	2.75	0.34	1-12	6.46-8.27
Comparison (37/3)	2.29	2.81	0.46	0-9	1.04-3.55
3rd Grade/Segregation					
Pretest					
Treatment (30/2)	1.10	1.12	0.21	0-4	0.52-1.68
Comparison (11/1)	2.27	1.56	0.47	0-5	0.79-3.76
Posttest					
Treatment (90/6)	5.27	3.77	0.40	0-14	4.22-6.31
Comparison (23/2)	2.47	1.34	0.28	0-5	1.69-3.27
4th Grade/Great Depression					
Pretest					
Treatment (24/2)	0.92	1.28	0.26	0-5	0.18-1.65
Comparison (18/1)	0.89	1.02	0.24	0-3	0.19-1.59
Posttest					
Treatment (65/5)	6.72	2.95	0.37	0-12	5.75-7.69
Comparison (30/2)	1.17	1.05	0.19	0-4	0.64-1.70
4th Grade/Progressive Era					
Pretest					
Treatment (60/4)	2.08	2.61	0.34	0-9	1.19-2.98
Comparison (27/2)	2.55	2.15	0.41	0-7	1.40-3.71
Posttest					
Treatment (82/6)	6.57	3.45	0.38	0-13	5.57-7.58
Comparison (48/3)	3.52	2.69	0.39	0-9	2.48-4.56
5th Grade/Native American History					
Pretest					
Treatment (42/3)	0.88	1.35	0.21	0-5	0.32-1.44
Comparison (22/1)	1.82	2.08	0.44	0-7	0.56-3.08
Posttest					
Treatment (85/7)	6.64	3.17	0.34	0-13	5.73-7.54
Comparison (40/2)	2.48	2.22	0.35	0-7	1.53-3.42
5th Grade/WWII Home Front					
Pretest					
Treatment (14/3)	1.21	2.78	0.74	0-9	0-3.45
Comparison (20/1)	0.20	0.52	0.12	0-2	0-0.53
Posttest					
Treatment (91/7)	6.45	3.25	0.34	0-14	5.55-7.35
Comparison (30/2)	0.57	1.40	0.26	0-7	0-1.27

*The number of items (k) was seven for all versions except 3rd grade Industrialization, where k=6. The maximum possible score for all tests is equal to 2k. n_s = number of students /n_c= number of classrooms.

In order to ensure that differences in scores on the narrative assessments could not be attributed to between group differences in general ability, two steps were taken. First, comparison schools were selected to match treatment schools based on their school-level median Core Total on

the Iowa Tests of Basic Skills (ITBS). Core Total scores include Reading Total, Language Total, and Mathematics Totals. The BHH II schools included three schools whose scores were in the middle two quartiles and one school in the bottom quartile, therefore the two comparison schools were recruited from among Iowa schools whose school level ITBS scores were in those quartiles, one from the bottom quartile and one from the middle two quartiles. Second, the actual ITBS scores of the students who took the narrative assessments were collected. Table 21 shows the mean narrative scores and mean ITBS Reading Total scores of students in treatment and comparison classrooms who completed the narrative assessments. The ITBS Reading Total was chosen as the score on which to compare performance because the children at one treatment school did not take the language tests so it was one of the only scores available and has a higher correlation with general ability than the Mathematics Total. As can be seen in Table 21, for all groups taking the assessments, the ITBS Reading Total scores for comparison students (expressed as National Percentile Ranks) were the same or higher than for treatment students, indicating that the students were similar in general ability. [The number of students and assessments in each treatment condition and grade level may differ slightly between Table 20 and Table 21. Table 21 includes scores only if we only had the complete data set, meaning for each student we had a narrative posttest (and pretest where relevant) and an ITBS Reading Total score while, Table 20 includes students for whom we had posttest (and pretest where relevant) whether or not we had ITBS scores.]

The ITBS data were collected not only as a means of showing that student groups were comparable, but also as a means of examining whether the BHH II Narrative Assessments measure something other than general ability, in this case, content knowledge of a particular history topic. As can be seen in Table 21, correlations between Narrative Scores and ITBS Reading Totals were not particularly high, indicating that Narrative Assessments appear to be measuring a construct other than general ability.

Table 21. Correlations of Narrative Assessment Scores with ITBS Reading Totals

Grade/Topic Test Condition	(n)	Narrative Total Mean (sd)	ITBS Reading Total (NPR) Mean (sd)	Correlation (*r*)
3rd Grade/Industrialization				
Pretest				
Treatment	(38)	1.68 (1.83)	63.66 (22.53)	0.03
Comparison	(22)	3.41(2.46)	79.5 (16.07)	0.43
Posttest				
Treatment	(56)	7.25 (2.67)	61.93 (24.19)	0.39
Comparison	(35)	2.37 (2.86)	70.63(20.79)	0.25
3rd Grade/Segregation				
Pretest				
Treatment	(21)	1.29 (1.23)	55.05 (25.89)	0.21
Comparison	(9)	2.00 (1.58)	78.89 (22.55)	0.30
Posttest				
Treatment	(59)	5.34 (3.67)	60.41(26.39)	0.36
Comparison	(20)	2.30(1.30)	68.05(23.78)	0.35
4th Grade/Great Depression				
Pretest				
Treatment	(19)	0.74(0.99)	70.42(21.61)	0.59
Comparison	(17)	0.94(1.03)	61.47(22.52)	0.33
Posttest				
Treatment	(59)	6.90 (2.99)	65.73(21.81)	0.24
Comparison	(27)	1.07(0.92)	64.52(24.03)	0.51
4th Grade/Progressive Era				
Pretest				
Treatment	(32)	1.78(2.66)	67.59(25.47)	0.29
Comparison	(18)	2.33(2.25)	75.89(19.01)	0.50
Posttest				
Treatment	(42)	6.29(3.54)	67.40(23.35)	0.47
Comparison	(34)	2.91(2.42)	70.03(20.94)	0.34
5th Grade/Native American History				
Pretest				
Treatment	(38)	0.82(1.29)	59.87(26.81)	0.00
Comparison	(20)	2.00(2.10)	68.15(27.4)	0.69
Posttest				
Treatment	(77)	6.77(3.15)	62.29(23.59)	0.34
Comparison	(37)	2.54(2.24)	68.46(25.08)	0.54
5th Grade/WWII Home Front				
Pretest				
Treatment	(14)	0.57(1.65)	67.74(24.34)	0.35
Comparison	(18)	0.22(0.55)	67.22 (27.21)	0.21
Posttest				
Treatment	(81)	6.69(3.22)	62.04(23.44)	0.30
Comparison	(27)	0.63(1.47)	69.04(24.77)	0.36

Sample Responses to Narrative Items

5th Grade

As can be seen in the tables above, the narrative assessments produced a wide range of responses from students. For the purpose of looking at the impact of learning history using the BHH curriculum, the following are examples of student responses. Some of the responses are student posttests paired with the same students' response to the pretest. Responses were

transcribed exactly as they were written with spelling and grammar errors intact and sometimes give an indication of the students' general writing ability. When describing responses, pseudonyms are used in all cases.

Figures 1 and 2 show examples of fifth grade students' pretest responses (before beginning the unit), followed by the same students' posttest responses after instruction. In both cases, the pretests received no points and the posttests were at least somewhat higher scoring responses. Anne's responses are shown in Figure 1. Students were not given specific instructions as to how to format their entries, only to "tell a story". On the pretest Anne chose only to say that she didn't know what the words meant, but in the posttest, she not only used the words, but demonstrated the facility of her use of them to make them part of a fictional diary entry, rather than just a list of definitions. Anne also expressed empathy in several instances during her narrative:

- By showing that the diarist "Jackie" would have understood that it was hard for her mother to have to go to work, saying it was, "…*the first smile I've seen in a while*"
- In her opinion of Japanese internment camps, "*Ever since Japan bombed Pearl Harbor everybody has been afraid of the Japanese so they sent them away, how stupid"*
- By showing that she understood not only what a Victory Garden was, but some reasons for growing one – to avoid using ration stamps and to save money so that they could "…*buy war bonds for the war effort*."

Figure 1. 5th Grade WWII narrative pre- and posttest responses from one treatment classroom student

Anne's pretest response:

Home front, Ration books, Victory garden, War Bonds, Internment camp, War effort and Rosie the riveter will have something to do with World War II, but I don't know what they mean.

Anne's posttest response:

Journal May 14th, 1942 Today my mom is staring work at a factory since most of the men aren't here on the homefront, but at war. My mom said women who started working were called Rosie the Riveter, so I called her that this morning and she gave me a smile, the first smile I've seen in a while! Since we got our ration books, we haven't been able to bake many things with a lot of sugar in them. My best friend was sent away to an internment camp because she is Japaness. Ever since Japan bombed Pearl Harbor everybody has been afraid of the Japanese so they sent them away, how stupid. Our principal has given talks about how we should plant Victory gardens so we don't have to use our ration stamps and to buy war bonds for the war effort, so I'm planning to do some of that this weekend. Well now that I've brought you up to date, I'll quit and go to my "Patriotic Duty" as my mom says and go gather the trash up!! Jackie

Figure 2 shows the responses of another student, Bob, who also scored low on the pretest and received a medium level score on the posttest, but still demonstrated his grasp of content and his empathy toward the people of the WWII home front. Bob chose to write his response in the form of a letter indicating with that choice that he understood that letters from home were a common way of communicating during WWII. Like Anne, Bob felt that the Japanese internment camps were "cruel" and showed his knowledge of the importance of the fictional "Rosie the Riveter" as a tool to encourage women to work.

Figure 2. 5th Grade WWII narrative pre- and posttest responses from another treatment classroom student

Bob's pretest response:

Home front -- don't know. Ration books -- to show what happened that day. Victory garden -- where one of the unions won the war. War bonds -- don't know. War effort -- they put all their effort into fighting. Rosie the riveter -- a woman that helped the soldiers fight.

Bob's posttest response:

Dear Matt, We're still on the Home Front. We have ration books nowe because product s are being rationed so there is more to send over to the troops. Because of the rationing people (including us) are planting Victory gardens so there is more food to send over to you. Mom is collecting stamps to try and get a war bond. Yesterday the government sent all the Japanese-Americans to Internment camps. I think it is cruel. Mom, Dad, and I are doing as much as we can to help the War Effort. Mom works with the Red Cross Volunteers now and I stay home and take care of dad. His broken leg is doing better. I guess that's sort of helping. "Rosie the Riveter" signs are up everywhere now to help encourage women to work in factories. Even though she is fictional, I really think she really is encouraging women. Write back soon. Love ya, Sally

The next response (shown in Figure 3) was a posttest from another student, John, in a different classroom. This response had numerous spelling and grammatical errors and omitted one term completely (war bonds), but still showed some of what John learned about the WWII home front. John made several points in his narrative that go beyond what was asked, such as:

- WWII was the start of daycare when children went to daycare so that their mothers could work
- Popular movie and TV stars helped advertise war bonds to encourage people to buy them
- Children and mothers were part of the war effort through rubber drives and saving bacon grease

Figure 3. 5th Grade WWII narrative posttest response from a treatment classroom student

John's posttest only response:

During World War II, on the home front Americans were afect by most of the men leving and most of the women started to fill mens work. That wich led to childred going to day cares or their parints would get a nanny to wach thim. War bonds were intedosd. Some of the populer tv stars would help sell them. SO not only could you buy one from a star, you would relly help the war! Food was rationed and Ration books were interdosed. They rationed food so they would be able to send it to the soiders. You would use the for food ather whis you woulde'nt have food. Victory gardens were gardens for people who were pourer then others Americans put Japanes Americans in Internment caps after the Japanes bomedPourl Haber. Every one trid to help with the war effort. Some children saved shose for rubber and mothers saved baccon grees for bombs for the war.

4th Grade

Figures 4 and 5 show examples of two fourth grade students' pretests before beginning the unit, followed by the same students' posttest responses after instruction. The pretests received no points and the posttests were higher scoring responses.

Susan's responses (Figure 4) demonstrate a great amount of learning. On the pretest, it was clear that Susan was trying to come up with reasonable definitions of things she really did not know anything about. On the posttest, Susan shows her knowledge about some details about the era:

- "…a lot of jazz and musical things going on" during the Roaring 20's
- The dust bowl happened at least in part because of, "..they kept planting wheat after wheat in the same spot."
- Many people blamed the Depression on President Hoover.

Figure 4. 4th Grade Great Depression narrative pre- and posttest responses from a treatment classroom student

Susan's pretest response:

The Great Depression is a war. A dust bowl is where there is a lot of dust. I think a penny auction is where there are a lot of people there for a sale. I think a migrant worker is a person that builds migrants. I think Hooverville is were people work ou live. Unemployment is that you are unmployed at a job.

Susan's posttest response:

During the Roaring 20s there was a lot of jazz and musical things going on, but when the 1930s came a long people called it The Great Depression because the stock market crashed and people lost their jobs and houses. The farmers had to go through the Dust Bowl, they got the dust bowl becuaes they kept planting wheat after wheat in the same spot so they got the Dust Bowl. When the stock market crashed farmers had to have Penny Auctions, it is when other people buy the farmers stuff for a penny and give it back to the farmer. A migrant worker is a migrant that works for famers. Some migrants lived in Hoovervilles because they lost there houses. They are call Hoovervilles becuase they blame Presidant Hoover. Unemployment is when you do not have a job, some migrants were unemployed when the stock market crashed.

In Figure 5, Mary's pretest shows that she already had some knowledge about the Dust Bowl (that it involved a drought), but not much else. Both of her responses have many spelling and grammatical errors, but she picked up quite a bit of knowledge about the Depression era (especially about geography) and her narrative included empathy and vivid imagery of the time ("…*many people could not even go outside without geting dust in your eyes*". *)*

Figure 5. 4th Grade Great Depression narrative pre- and posttest responses from a different treatment classroom student

Mary's pretest response:

I know some body that live in the Roaring 20's. She also lived in the Great Depression. That was when money was hard to get. During that time a dust bowl was going on. That was when a huge drought went on. She did not like Penny auctions. A penny auction is when you bet pennys or dollers. Migrite worker was her dad. She lived near Hovorville. Her dad could not be a unemployment worker.

Mary's posttest response:

in the roaring 20's were really good people made money quickly. In the great depression people had no money, and some even lived in a hoovervill, a hoovervill is were people who did not have a house lived. Munny people did not have jobs, this was called unemployment. Most hoboes were Migrant workers, that means to move around to maker monery ey: some hobo works in organ for the apple havest then moves down to Gorga for the peach havest. The dust bowl was really bad expespely in Kanses, Colorado, Oklahoma, and Texas. All the farmers crops died and Big dust storms covered the land and many people could not even go outside with out geting dust in your eyes. But becouse they could not plant cropes the had penny auctions penny auctions were when people who lost there farms and people from town would make one low bid then on more people would bid.

Figure 6 shows another example of a somewhat high scoring posttest response. Tom appears to have tried to write his entry in a way that connected many of the terms in causal relationships.

Figure 6. 4th Grade Great Depression narrative posttest response from an additional treatment classroom student

Tom's posttest only response:

In the 1920s everyone was having a good time with all their money. But when the Stock Market Crashed things got messy. Hobos started making Hoovervilles to live in when looking for jobs. There was a lot of unemployment. Migrant workers where running west to California looking for jobs. When farmers lost their money and couldn't pay back the bank for the loan they used thats where a Penny Auction came in. The back was selling farmers stuff for one penny. When farmers planted the same crop year after year the Dust Bowl started. Big clouds of dust came, and destroyed everything a farmer had and when that happed they became a Migrant. Thats the Great Depression.

3rd Grade

Figures 7 and 8 show examples of the pretests of two third grade students before beginning the unit, followed by the same students' posttest responses after instruction. The pretests received zero or few points and the posttests were high scoring responses. In Figure 7, Jacob responded in strong language about how wrong he believed slavery was, but it was the only term he used. By the posttest, Jacob appeared to be still as angered by the idea of slavery, but also understood more about why it existed in the US. He understood that segregation replaced slavery and that it continued to have a negative affect on the lives of African Americans.

Figure 7. 3rd Grade Segregation and Slavery narrative pre- and posttest responses from a student in a treatment classroom

Jacob's pretest response:

Slavery is when the south is very, very, very mean to the people. They beat them, they make themw ork really, really, really hard and their just to mean to them. That means they belive in hurting people that didn't do anything to them. They also have live in a very small house. I do not belive in slavery at all. I think they sould be nice to them. They are just as special as any other person. That is what slavery is.

Jacob's posttest response:

Slavery is when African Americans blacks had to work and not get paid. If they did stuff wrong they'd get whipped by their Masters. The Constitution allowed all of this to happen. The Constitution is a papper that tells us what we can do and what we can't do. Soon after many years there was a change in a rule which was call an Amendment. Then after the Amendment there was a 13th Amendment. That ment that Slavery did no longer exist anymore. Segregation ment that blacks and whites couldn't be equal. They had to be separated. Whites from the south were Prejudice because slaves were free. Most whites where mean to the blacks. Their were Jim Crow Laws that ment blacks couldn't go to the same places as whites and couldn't drink out of the same drinking fountains.

Another 3rd grade student, Kate (shown in Figure 8), also progressed from a fairly shallow knowledge of what slavery is (and that Martin Luther King, Jr. was somehow involved) to understanding quite a bit about the Constitution, amendments, and segregation. Kate appears engaged with the material and motivated to tell as much as she has learned, some more relevant than others.

Figure 8. 3rd Grade Segregation and Slavery narrative pre- and posttest responses from a different student in a treatment classroom

Kate's pretest response:

Slavery is when peple treat other people bad. The slaves are usally blacks. The blacks were treated porely. They got whipped and hit. Segregation is when the blacks were treated bad. MLK Jr. was a very famous black. He made a lot of speches. His most famous words were, "I had a dream.

Kate's posttest response:

Slavery was an awful thing. Slavery was when white people owned blacks. Slaves were brought over from Africa. Harriet Tubman was a formen slave. She helped over 300 people to the North. The Constitution is the laws that are made. The whole country has to follow In different countrys they don't have presedents. Our presedent is George W. Bush. An Amendment is the change in a rule. When the slaves crossed the Mason Dixon line there was a rule that passed. The slaves had to go to Canada to be free. the 13th Amendment is when slavery ended. It ended in 1865. It was a very good time for the African Americans. The blacks hated it, Abraham Lincoln wrote the Emancipation Proclamation. It freed the slaves. The word Segregation means seperate. The blacks hda to do everything seperate than the whites. Jackie Robinson was a black. He played baseball. He was the first even black to paly baseball. he was my favorite black player. The word prejudice means to criticize a person. The other person does not even do anything. That is what the whites did to the blacks. I feel sorry for the blacks. The Jim Crow laws were awful. They were laws that blacks had to follow. some of them were do going in the same restrooms. On only sitting in the back seat.

Andrew's entry (Figure 9) did not get a high score, because he did not use all the terms but is of note because of his learning of somewhat rote ideas, like that prejudice means to "pre-judge" and his closing statements, "There is still segregation. It is not FAIR!"

Figure 9. 3rd Grade Segregation and Slavery narrative posttest response from a student in a treatment classroom

Andrew's posttest only response:

I think that segregation is interesting. It is when someone else does not like a person for who they are on the outside. That can also be called prejudice, when you "pre-judge" someone. People made jim crow laws like blacks could not drink at whites only water fountains, or sit on the main floor of a theader or swim in whites pools and so on. Segregation was after the 13th amendment, but before that there was slavery. That is when a white owns a black like properdy. That stopped when they made an amendment to The Constitution. The 13th, There is still segregation. It is not FAIR!

5.2.3 Photo Analysis Assessments

A random sample of students (at the classroom level) in 3rd through 5th grades completed written photo analysis assessments concerning the six *BHH* content areas implemented as one of three types of written assessments intended as direct means of examining student learning outcomes that occurred as a result of the *BHH II* project. Assessments asked students to examine a photo that was relevant to the content area they were studying and answer several questions concerning the photo: "What do you think is happening in the photo?", "Who do you think the people in the photo are?", "When do you think the photo was taken?", "Where do you think the photo was taken?", "Why do you think the photo was taken?", and "After you look at the photo, what questions do you still have about the photo that you would like to learn the answers to?" For this analysis, we looked at students' answers to the "What", "Who", "When" and "Where" questions. Resulting analyses were scored with 0, 1, or 2, points possible for each of the four questions, so the maximum score possible on all photo analysis assessments was eight. Estimates of reliability of these scores in the form of Cronbach's alpha for each grade level and content area are provided in Table 8 in the Methods section of this report.

Treatment and comparison classrooms in which students completed photo analysis assessments were assigned to either a *pre-post* or *post only* condition. Approximately half of the classrooms completed both a pre and a posttest (identical test for pre and post administration) and the other half completed the posttest only. Scores on all posttests were pooled for this analysis. [The effects of the testing condition (pre-post vs. post only) were minor are not included in this report.] The effects of treatment condition on student performance on the photo analysis measures are summarized in Table 22. Data are reported at the student level. Future research will include analyses of classroom level data.

Table 22 presents the important comparisons between treatment students on pre and posttests and between treatment and comparison students. The table shows means for pre and posttest performance along with standard deviations, standard errors, range of scores, and 99% confidence intervals for each test and treatment condition. Comparison of group means between two groups with non-overlapping intervals indicate 99% likelihood that the difference between the two means is not due to chance.

As can be seen in Table 22, students in treatment schools on average scored higher on photo analysis posttests than they did on pretests indicating that growth in ability to use photos to talk about historical content knowledge occurred over the course of the project. In addition, students in treatment schools scored higher on the photo analysis posttest than did comparison

school students over the same school year, indicating that the historical content and/or analysis skills were not something that children of the same grade typically learned during the school year. Treatment students and comparison students did not differ on pretest scores.

Increases in treatment students' scores from pretest to posttest occurred at all grade levels and for all units, with increases in group means ranging from 1.3 to 3.7. The gain in mean scores for the 5th Grade Native American history unit was not as large, with slightly overlapping 99% confidence intervals for pre and posttests. Students in general scored low on this assessment with a treatment student posttest mean total of 2.26, which was much lower than the other treatment group posttest means.

Among comparison students, group means changed between pre and posttests, varying in amount and direction from -0.4 to +1.27. The largest comparison group gain of 1.27 was on the 3rd segregation photo analysis on which treatment students' gain between pre and posttest means was 3.5. All differences between treatment students' mean pre and posttest mean scores (with the exception of the 5th grade Native American History scores mentioned above) were significant as illustrated by non-overlapping 99% confidence intervals. Using the same criteria, all differences between treatment and comparison students' mean posttest scores (with the exception of the 3rd grade industrialization scores, which also had slightly overlapping 99% confidence intervals) were also statistically significant.

Table 22. Comparisons of Pre/Post and Treatment/Comparison Conditions for Student Photo Analysis Assessments

Grade/topic Condition (n_s/n_c)	Mean (k=4)*	SD	SE	Range	99% Confidence Interval
3rd Grade/Industrialization					
Pretest					
Treatment (55/3)	1.78	2.72	0.37	0-8	0.81-2.76
Comparison (11/1)	1.45	2.13	0.65	0-6	0-3.52
Posttest					
Treatment (104/6)	4.04	2.20	0.22	0-8	3.48-4.62
Comparison (23/3)	2.56	2.59	0.54	0-6	1.04-4.08
3rd Grade/Segregation					
Pretest					
Treatment (27/4)	0.48	1.74	0.33	0-7	0-1.41
Comparison (11/1)	0	0	0	0-0	n/a
Posttest					
Treatment (58/7)	3.98	3.20	0.42	0-8	2.87-5.10
Comparison (37/2)	1.27	2.49	0.41	0-8	0.16-2.38
4th Grade/Great Depression					
Pretest					
Treatment (58/4)	0.60	1.56	0.20	0-5	0.06-1.15
Comparison (31/2)	0	0	0	0-0	n/a
Posttest					
Treatment (63/5)	4.30	3.09	0.39	0-8	3.27-5.34
Comparison (45/3)	0.98	2.04	0.30	0-7	0.16-1.80
4th Grade/Progressive Era					
Pretest					
Treatment (37/3)	2.40	2.81	0.46	0-8	1.15-3.66
Comparison (11/1)	2.00	2.93	0.88	0-8	0-4.80
Posttest					
Treatment (78/6)	5.83	2.01	0.23	0-8	5.23-6.43
Comparison (40/3)	1.63	2.40	0.38	0-6	0.60-2.65
5th Grade/Native American History					
Pretest					
Treatment (30/3)	1.00	1.68	0.31	0-6	0.15-1.85
Comparison (10/1)	0	0	0	0-0	n/a
Posttest					
Treatment (47/6)	2.26	2.80	0.41	0-8	1.16-3.35
Comparison (39/3)	0.28	1.02	0.16	0-5	0-0.72
5th Grade WWII Home Front					
Pretest					
Treatment (54/5)	1.57	1.99	0.27	0-5	0.85-2.30
Comparison (0/0)	n/a	n/a	n/a	n/a	n/a
Posttest					
Treatment (90/7)	4.26	2.66	0.28	0-8	3.52-4.99
Comparison (23/1)	0.61	1.37	0.28	0-4	0-1.42

*The maximum possible score for all tests was equal to 2k, or 8 for all photo analysis assessments.
n_s = number of students /$n_{c=}$ number of classrooms.

As with the narrative assessments, in order to ensure that differences in scores on the photo analysis assessments could not be attributed to between group differences in general ability, two steps were taken. First, comparison schools were selected to match treatment schools based on their school-level median Core Total on the Iowa Tests of Basic Skills (ITBS). Core Total scores

include Reading Total, Language Total, and Mathematics Totals. The BHH II schools included three schools whose scores were in the middle two quartiles and one school in the bottom quartile, therefore the two comparison schools were recruited from among Iowa schools whose school level ITBS scores were in those quartiles, one from the bottom quartile and one from the middle two quartiles. Second, the actual ITBS scores of the students who took the photo analysis assessments were collected. Table 23 shows the mean photo analysis scores and mean ITBS Reading Total scores of students in treatment and comparison classrooms who completed the photo analysis assessments. The ITBS Reading Total was chosen as the score on which to compare performance because the children at one treatment school did not take the language tests so it was one of the only scores available and has a higher correlation with general ability than the Mathematics Total. As can be seen in Table 23, for all groups taking the assessments, the ITBS Reading Total scores for comparison students (expressed as National Percentile Ranks) were the same or higher than for treatment students, indicating that the students were similar in general ability. [The number of students and assessments in each treatment condition and grade level may differ slightly between Table 22 and Table 23. Table 23 includes scores only if we only had the complete data set, meaning for each student we had a photo analysis posttest (and pretest where relevant) and an ITBS Reading Total score while, Table 22 includes students for whom we had posttest (and pretest where relevant) whether or not we had ITBS scores.]

The ITBS data were collected not only as a means of showing that student groups were comparable, but also as a means of examining whether the BHH 2 Photo Analysis Assessments measure something other than general ability, in this case, content knowledge and skill development of a particular history topic and photo analysis skills. As can be seen in Table 23, correlations between Photo Analysis Scores and ITBS Reading Totals were not particularly high, indicating that Photo Analysis Assessments appear to be measuring a construct other than general ability.

Table 23. Correlations of Photo Analysis Assessment Scores with ITBS Reading Totals

Grade/Topic Test / Test condition (n)	Photo Analysis Total Mean (sd)	ITBS Reading Total (NPR) Mean (sd)	Correlation (*r*)
3rd Grade/Industrialization			
Pretest			
Treatment (32)	1.16(2.27)	58.88(25.23)	0.19
Comparison (10)	1.60(2.22)	60.90(22.92)	0.54
Posttest			
Treatment (55)	4.15(2.33)	59.71(26.15)	0.49
Comparison (21)	2.28(2.53	68.76(23.41)	0.23
3rd Grade/Segregation			
Pretest			
Treatment (18)	0.33(1.41)	60.28(26.27)	0.26
Comparison (9)	0(0)	57.22(9.16)	n/a
Posttest			
Treatment (27)	3.22(3.41)	57.59(24.74)	0.38
Comparison (31)	1.16(2.45)	70.61(17.42)	0.44
4th Grade/Great Depression			
Pretest			
Treatment (30)	0.80(1.83)	65.33(22.85)	0.18
Comparison (27)	0 (0)	66.19(20.78)	n/a
Posttest			
Treatment (37)	5.11(2.84)	69.73(19.48)	0.35
Comparison (37)	1.05(2.10)	70.95(21.36)	-0.04
4th Grade/Progressive Era			
Pretest			
Treatment (28)	2.00(2.62)	66.82(22.14)	0.23
Comparison (8)	1.50 (2.98)	76.25 (19.90)	0.33
Posttest			
Treatment (52)	5.44(2.04)	71.31(19.24)	0.35
Comparison (35)	1.74 (2.48)	66.74 (24.33)	0.33
5th Grade/Native American History			
Pretest			
Treatment (28)	0.96(1.69)	68.07(20.63)	0.28
Comparison (9)	0(0)	71.79(20.32)	n/a
Posttest			
Treatment (44)	2.41(2.83)	69.82(20.70)	0.35
Comparison (29)	0.38(1.18)	67.93(24.32)	0.25
5th Grade/WWII Home front			
Pretest			
Treatment (47)	1.66(2.02)	59.74(24.00)	0.23
Comparison (0)	n/a	n/a	n/a
Posttest			
Treatment (49)	4.42(2.58)	58.47(25.64)	0.34
Comparison (20)	0.70(1.45)	69.10(26.77)	0.48

Sample Responses to Photo Analysis Items

Responses to the photo analysis questions were typically quite a bit briefer than were the responses to the narrative items. Students' entire responses to all questions, "Who", "What", "When", "Where" and "Why" and "What other questions do you still have about the photo?" were used to arrive at each of the separate scores. Therefore, they provide less clear-cut illustrations of students' content knowledge acquisition than do students' responses to narrative prompts, but a few

examples demonstrate how students' ability to make sense of the photograph based on content knowledge and acquisition of photo analysis skills changed over the course of the unit implementation. Responses to the photo analysis tasks also provide information about the nature of the content students learned and how students learned to use photographs to make inferences about history using primary sources. For many students, the photograph acted as a stimulus to tell other things they knew about the topic. In the sections below, actual responses are given, but the name associated with each response is a pseudonym.

5th Grade

For the WWII home front unit, students examined a 1942 photograph taken in San Francisco of several Japanese-Americans en route to internment camps.

The responses given by a 5th grade student to the questions concerning the photo at both pre and posttest are shown in Figure 10. Although from these responses, it is difficult to tell the depth of Daniel's knowledge concerning Japanese internment, it is clear that he has learned some basic points – that the people were Japanese-Americans, that it was during WWII, happened in the US, and that this was a hardship for the Japanese-American people.

Figure 10. 5th Grade WWII home front photo analysis pre- and posttest responses from a treatment classroom student.

Question	Daniel's pretest responses	Daniel's posttest responses
What?	I think people are leaving this city	The Japanese are getting moved to Recolection camps
Who?	People that are working or leaving the city	Japanese American and Army men
When?	I think this picture was taken not to long ago	1944
Where?	I think this photo was taken in the past	In USA
Why?	To show how people lived long ago	So people know the hardship of the Japanese internment

Another 5th grader**,** Jane, whose responses are shown in Figure 11, also demonstrated her knowledge concerning the bombing of Pearl Harbor and why Americans thought Japanese-Americans needed to be in camps.

Figure 11. 5th Grade WWII home front photo analysis pre- and posttest responses from another treatment classroom student.

Question	Jane's pretest responses	Jane's posttest responses
What?	People that are waiting for a train	Japanese-Americans are being sent to camps because they think they might give out secrets to the Japanese.
Who?	Business people going to an important place	I think the people are Japanese Americans.
When?	Way back then around maybe	Around the time when the Japanese bombed Pearl

	1600s	Harbor.
Where?	Maybe a bus stop	At a train station when they were leaving to camps.
Why?	They could show what it would look like if you were at a bus stop	To show where they were going.

4th Grade

The photograph used for the 4th grade photo analysis was a photo of a Hooverville camp taken by Dorothea Lange. The assessment included a half-page size photo and did not include credits of any type. As reflected by the range of scores students received for their responses to this assessment, the quality of responses varied greatly to this photo. Figures 12, 13 and 14 show some examples of somewhat strong student responses. Figure 12, shows a high scoring response given by William who managed to include information about migrant workers, the stock market and the dust bowl into his response to a posttest only condition. He also displayed empathy for the people in the Hooverville by saying it was "painfull" (sic).

Figure 12. 4th Grade Great Depression photo analysis posttest response from a treatment classroom student.

Question	William's posttest only responses
What?	The people are homless because they are poor and got there house taken away and all of there stuff with there job they are looking for one
Who?	Migrant worker looking for a job and a home to live in.
When?	In the 1930's when people were poor becaue the stock fell and they were hit by the dust storm in the US Dust Boul.
Where?	In the southern US
Why?	To show you about the 30's and how it was and how painfull.

The responses given by a 4th grade student, Sarah, to the questions concerning the photo at both pre and posttest are shown in Figure 13. Although these responses are very simple and brief, in the posttest she tells that the people are poor, unemployed, and uses the correct term, Hooverville, and when paired with her pretest, it is clear that the content knowledge is new to her.

Figure 13. 4th Grade Great Depression photo analysis pre- and posttest responses from another treatment classroom student.

Question	Sarah's pretest responses	Sarah's posttest responses
What?	I think they are building a town.	I think these people are looking for work.
Who?	People building a town	I think these people are poor people.
When?	1857	I think this photo was taken in the 1930s.
Where?	Iowa	I think it was taken in Hooverville.
Why?	To remember this in the future	I think someone took this photograph to show us what it was like during the depression.

Paul's responses (Figure 14) are somewhat similar to Sarah's and again illustrate that the students' responses do not need to be lengthy to show that they know quite a bit about the content.

Figure 14. 4th Grade Great Depression photo analysis pre- and posttest responses from an additional treatment classroom student.

Question	Paul's pretest responses	Paul's posttest responses
What?	I think people are looking at graves.	People are setting up camp.
Who?	I think they are solidgers.	Hobos
When?	1856	In the 1930s
Where?	I think it was taken at a battle field.	In a Hoover Ville.
Why?	To show what it was like after a war.	To show what life was like back then.

3rd Grade

For the 3rd grade Segregation unit photo analysis assessment, students analyzed a photograph taken in 1938 on a South Carolina courthouse lawn of an African American boy drinking from a fountain labeled "colored"(Figure 15). As in all the photo analysis assessments, the students had a larger version of the same photo, with no credits or other information supplied. Because 3rd grade writing in general is typically not as well-formed as the older students, sometimes the responses were difficult to interpret, but the examples below show how it was still possible to look at growth in knowledge and skill over time.

Figure 15. Photograph used for 3rd grade Segregation photo analysis assessment

Source: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA, Reproduction No: LC-USZ62-100414

Figure 16 shows Andrea's responses to the pre and posttests. Her spelling and grammar give indications that she is not probably not of high general ability, yet she exhibits some knowledge of the segregation era, the relevance of the "colored" sign, and that this photograph has something to do with the "sivl rits"(sic – civil rights) issue.

Figure 16. 3rd Grade Segregation photo analysis pre- and posttest responses from a treatment classroom student.

Question	Andrea's pretest responses	Andrea's posttest responses
What?	(blank)	A little cored girle is brinking out of a colred brinking fownten.
Who?	Slav	I do not know
When?	log ang	1866
Where?	otesild	In the soth
Why?	fon cos.	They whoted to show us the sivl rits
Other questions?	Why bo's it sald colored	nuthing

On the other hand, in Figure 17, Adam seems to know something about segregation even before the unit begins, indicated by his question of "Why were there laws about black and white people against each other?" However, his answers to the questions on the posttest show a new sophistication in understanding of the problem, in his inclusion of "Jim Crow Laws" and his statement that the laws were "unfair".

Figure 17. 3rd Grade Segregation photo analysis pre- and posttest responses from another treatment classroom student.

Question	Adam's pretest responses	Adam's posttest responses
What?	A boy is drinkin	A person is getting a drink from a "colored only" water fountain
Who?	A colored boy maybe a slave	I think it is a little boy.
When?	Around 1953	Aroud the 1950's and 1960's, during the Jim Crow Laws.
Where?	Possibly Atlanta GA, Pennsalvania	Probly down south of the United States of America
Why?	To show the future what it was like then.	Probly to show how the Jim Crow Laws were unfair to a lot of peple.
Other questions?	Why there were laws about black and white people against each other.	Why did people start Jim Crow Laws? Who started Jim Crow Laws? When did Jim Crows Laws start?

Jenna's responses (Figure 18) on the posttest shows careful attention to many details in her descriptions of the photo while still keyed in to the central issue, blacks and whites had separate drinking fountains.

Figure 18. 3rd Grade Segregation photo analysis posttest only response from a treatment classroom student.

Question	Jenna's posttest only responses
What?	a boy is drinking at a colored water fountain.
Who?	He is a black Afrcain Amairacan boy he is wearing a shirt, pants, and shoes. He is probably around 9 years old. He looks like a nice person.
When?	Maybe somewhere around 1952 or somewhere close to that.
Where?	Maybe in Alabama at a colored water fountain.
Why?	To show people what black water fountains looked like and how they were seperated from the whites water fountains.
Other questions?	Why did they make black people use different thining that the white people used

Other Photo Analysis Data

In addition to the written assessments employing photo analysis, some students also analyzed photographs during group interviews. The two main purposes of analyzing photographs in a group setting were to: 1) allow students a chance to talk about the photos without the constraints of writing, and 2) examine whether working in a group would reveal more about the depth of students' knowledge about topics. At least one group at each treatment school participated in interviews and one group at each grade level participated in group interviews at one of the comparison schools. Summaries of group interviews at each grade level are included in Appendices of this report along with complete redacted transcripts of each group interview. Some relevant results from the summaries are presented here, by grade level.

5th Grade

Fifth grade students in the group interviews had several opportunities to respond to photographs. The photograph used was taken in the early 1940s and depicts several women working inside an airplane fuselage. As with the written photograph analysis tasks, the photographs used to do this task students were larger, approximately 6" by 8".

The format of the group interviews was somewhat informal, but the types of questions raised by the interviewer were similar to the written photo analyses, usually including, "What do you think is happening?" "When do you think the photo was taken?" "Why might it have been taken?" and "How do you think the people in the photograph might have been feeling?" With the treatment school 5th graders, it was apparent that the group format contributed to correct inferences about the photo. A couple of the groups started off on the wrong track by one child saying it [the airplane] looked like a spaceship, but when another child mentioned the correct idea, the others quickly followed into the discussion based on more accurate inferences. Some students' statements that exhibited the richness of their knowledge included:

- "I think they're women trying to help the war effort by working – in factories. It looks like they're inside a plane there, working on maybe the riveting or welding something."
- "All the men were at war and there was, like there was only women left; and back then women didn't wear jeans yet and the women here wearing jeans and they're working."

When asked how the women might be feeling, the students continued to demonstrate knowledge of the times. Their responses included:

- "Glad to be working for, like to be working in factories for the first time and just helping out the war effort."
- "They might be like tired or like they might be thinking how – why they're doing – helping because their husbands are away and they feel sad."
- "Pretty important or excited that they get to help with World War II."

Fifth grade students who took part in group interviews also did a different type of photo analysis task that involved looking at four different photos taken during the time of Japanese Internment camps and then telling a story about the photos. The photos were numbered in sequential order and students had to tell the story of what was happening in the photos using the sequence. The four photos that the students viewed depicted, 1) a well-dressed, happy looking Japanese American family on the front lawn of a middle-class home, 2) a sign in the window of a Japanese-American owned business saying that they were closing the business, 3) a group of Japanese-Americans waiting to board a bus, suitcases in hand, and 4) a photo taken at Manzanar, one of the internment camps.

Students in the treatment group schools told a plausible story to go with this sequence of photographs and the positive influence of other knowledgeable group members was apparent. They immediately identified the story as that of Japanese Americans in the US during WWII. In one group, the discussion started like this:

Child: During World War II and I think it's where the Japanese internment -- where the Japanese Americans got kicked out of their homes and had to go to reclo- recol-

Another child: Relocation!

Child: Relocation camps.

Interviewer: Okay, anything else you want to add? Why do they have to go to in to internment camps?

Child: Probably because they blamed it on the Japs that why they bombed America and Pearl Harbor and stuff.

[It should be noted that students learned parts of what they knew about internment camps from primary sources including newspaper articles where the term "Japs" was used. Classrooms teachers discussed the use of racially offensive epithets, however this did not prevent students from using this term in this context.]

In a group interview at another treatment school, the students began their responses to the photos by saying:

> Student 1: The family was like a Japanese family and they got taken, well they had to go, and go to a bus stop and get ready to leave for Internment Camp. And then they got brought to them and that's were they had to live in the Internment Camp.
>
> Student 2: In—this picture [photo 4] looks like it's in a Japanese Internment Camp homes. It looks like it's out in the desert, by the dead ground and the grass and the kind of sandy look. It's kind of…well, there's mountains in the background, but you can see people walking around in the fenced in areas.
>
> Student 3: Like if Japanese Americans for—say they owned a store and when Pearl Harbor was bombed and the Japanese bombed it the Americans didn't want them in American anymore so, they would close their stores and kick them out of their houses. And then the Americans threw up these houses and then they would send the Japanese Americans to them. And some would come after the war, but some were sick and died and when they would get home some of there stuff would be destroyed or gone. If they owned a store, it might have been destroyed or something.

Students' empathetic response to and depth of knowledge concerning the Japanese internees was evident in responses like these:

- "The Japanese that are leaving and they put a note on their store. And in the second picture they have a picture of their home in the, and all the family; and I think they took that – I guess that the last picture they took because, because they were going to – they were leaving somewhere else and they didn't know if they were going to live in the same house, or what was going to happen to other people in the house. In the third picture I see the people packed up and they could only take one or two suitcases with them, and I see a little kid crying to leave his home, and the rest are getting on the bus. And the third picture, I see the camps in the desert, and I think it would've been horrible for them and kind of, they got used to it because they were starting to do some gardens and they were opening schools and

they started to think about 80 some baseball teams, they had 80 some baseball teams that they started there."

- "And when they went to the Internment Camps most Americans supported them, so they probably—when they came back they probably felt like—we probably felt like they were, like, being mean to the Japanese by not giving them a fair chance."
- "It must have been really, really, really bad for the Japanese to be at the Internment Camps because they were forced out of their homes by people who thought that they were going to like help the Japanese and then they weren't allowed to have very many of their possessions, they could only take the stuff that was necessary and they could only take two suitcases, so it must have been really bad for them."
- "They're sad and everything and the Japanese Internment Camps were like bad because there was no insulation and it was like hard in the winter. Cause it was always cold."

In contrast, comparison group students were unable to offer correct inferences about the series of photographs that would help them construct a narrative. One student in the group said that he thought they were Japanese people, but when probed about why he thought so said, "'Cause their eyes look weird" and made a gesture to demonstrate their eyes. The comparison groups did little to interpret the photos, making only statements such as, "this one shows them getting into…a bus" and "...in a desert by the mountains."

4th Grade

Fourth grade students also took part in group interviews that included several different opportunities to discuss photographs from the Depression era. As with the 5th graders, group dynamics appeared to contribute to the students' ability to use the photos to reveal fairly deep relevant knowledge about the Great Depression. For example, students were shown a photo taken in the mid 1930's outside a New York employment agency. It depicts a crowd of people searching the job listings.

The fourth grade students, in general, had a harder time than the fifth graders staying focused on the task that we were asking of them, but their discussion was not idle, mostly comprising talk about other things they knew about the Great Depression. Two groups pretty much completely were unable to identify anything about the actual setting of this photo, instead deciding it had something to do with reading stock market quotations and talking at length about the stock market and another group used the photo merely as a jumping off point to talk at length about "alphabet agencies". Their discussion included:

Student: Alphabet agencies are where…President Roosevelt when he became President everybody lost…some people lost their jobs and he wanted them to work so they could get money for their family so they didn't have to live in Hoovervilles, so then he did the alphabet agencies.

Another student: Alphabet agencies means that when President Roosevelt – he made these, like new agencies that the governments can help the people. But he didn't want them just hand out the money. He just wanted to give them a job and then they can start doing their own things.

A third student: The alphabet agency is President Roosevelt just didn't want what President Hoover was doing which is just making them sit back and he had the government give them money. President Roosevelt made them… made little… made alphabet agencies like he named them CCC and the W…WAC or something. They all had letters, like, in them. And they were little jobs. Like somebody can help build a park or something.

Students in the 4th grade interview groups also viewed a photograph of the Dust Bowl. This photograph was one of the very few that was also recognized by the comparison student group. The Dust Bowl is often discussed in a typical 4th grade curriculum unit about the Southwest region of the U.S. However, the depth of the discussion was very different with only one comparison student saying that it was the Dust Bowl which was kind of like a drought. The comparison students were unable to place it correctly in time and only one student associated the Dust Bowl with the depression. In contrast, the treatment groups provided many details, including the following:

- It happened in Oklahoma and states near there
- It was caused by drought and overcultivating land with severe loss of topsoil
- There was danger to humans and to livestock from dust and lack of food for animals
- Additional damage was done by grasshoppers
- 35 million acres were destroyed for cultivation
- There were "Black Blizzards"

One group's discussion went like this:

Student: Like they got done planting it over and over and over again and then the soil staring to like, loosen up and then when the wind blew it made this big cloud of dust, so if we were to roll over a house, it be all dusty.

Student: The people, it was more, the Dust Bowl really started in Oklahoma and then moved to the other states around it. So everybody in the, like in the mid-west were being hit by a whole bunch of dust and they had to cover their faces with, like, washcloths and stuff like that to keep the dust out of their mouths.

Student: Well, it costs lots of money because it destroyed everything and then, they had to get the farms back and, well had to travel to California I think or something.

Student: They had to like, cover up food and cover up milk and water and their mouths so when they go outside they don't get covered in dust.

Student: They took – when the Dust Bowl hit they took rugs and put them over their windows and doors so that dust couldn't get in.

3rd Grade

The third grade group interviews were slightly different from the ones with the older students because it is quite common for students in all classrooms to study segregation and civil rights in some form during 3rd grade. Therefore, comparison group students were more able to engage in conversation about the photographs, although like their 4th grade counterparts, their conversation was less detailed and more surface than that of their treatment group peers. For example, the first photograph shown to the two groups was of an African American man waiting at a North Carolina bus station, near a sign that read "Colored Waiting Room", taken in 1940.

In this case, all groups, both treatment and comparison knew something about segregation but the treatment students tended to show evidence of deeper and more detailed understanding of the times, using appropriate vocabulary and correctly placing the photograph in time. For example, one group's discussion included:

One student said that maybe the man in the photo was waiting because he had to go on the train last, adding, "Because they're black – because it's prejudice and because white people thought they were better than blacks." This was followed up by these comments:

Student: Because the whites are trying to keep blacks down so that they can stay up.

Interviewer: Okay, and what does that mean?

Student: It means that they're trying to keep the blacks, like they're still the slaves even though they had the rules that had been passed.

Student: Because of the Jim Crow laws.

Interviewer: What are the Jim Crow laws?

Student: They're laws that keep blacks down and waiting longer.

Interviewer: and waiting longer? What other kinds of things do the…

Student: They can't do, like much. And one of them, they have to take a test and pay a tax.

Student: Because of segregation and because it's pre-judging people just because of the color of their skin.

The comparison group students knew that the photograph depicted segregation but were less able to come up with relevant discussion. They had some general accurate ideas about what segregation meant, but did not have extensive knowledge and the comparison group students were quite confused as to when the photograph might have been taken. After the question was first asked, one comparison student said, "It looks kind of like Martin Luther King." Another student said, "He looks like he was going in to some kind of place." Another student said, "It looks like the man, he must be colored, and he has to wait, because he's colored and probably then they tried getting all the white people on the buses or on the train first, and then if all the white people were gone there are no more, they'd try getting black people on the back." The other students did not add details to this idea and one student said, "It looks like he's going to go in the train and start shooting people."

This discussion continued in a scattered fashion as above with one comparison student saying, "He looks like he's kind of mad and it looks like he's kind of sad, and it looks like he's kind of glad, and it looks like he's kind of everything. He might be glad because he might go there or he might be sad because of the colored waiting room, and he might be mad because he has to wait for the train." Another comparison student said, "He's kind of mad because he's, well, he's like a different color and it's not fair that he don't get to talk to white people or, like, be friends with them or any of that stuff." The comparison group did not name the laws as being Jim Crow laws or talk about people who were involved (other than mentioning "Martin Luther King") and their time estimates varied from 1977 ("because only black people could drink out of certain water fountains and stuff") to 1700s (because of the black and white photo) to 1856 (without a reason). One person said it was when Martin Luther King was alive.

The second photograph that students examined during the group interviews showed voting rights protestors in Mississippi during Freedom Summer in Mississippi in 1964.

Students' reactions to this photograph differed greatly between treatment and comparison groups. For the most part, treatment students were able to establish a correct context and deep

understanding of the issues of voting rights violations for African American voters, while the comparison group was entirely unable to establish a plausible context for this photograph. Some comments made by treatment students included:

- "Black people want to vote, but whites keep making more laws because of the 13th amendment. They said that they could vote, but they had to pass a test, a reading and writing test, and a tax, a liberty tax."
- "They didn't want to go because of the violence, they could hang them, they could throw rocks at them, they could beat them up, they could shoot them."
- "They wanted to end the literacy test and since the 15th Amendment says that they can vote, they still had the violence and they were afraid of that and they were afraid of – they didn't want to have to do the literacy test because they don't know how to read or write. They don't know how to write their name or anything."
- "I think those two girls with the signs, I think they're being really prejudiced" but another student challenged the first student and said, "I think they want them to vote…they're standing with them and they're sticking up for them and themselves."

In general, the treatment groups were able to understand the problem depicted, sometimes showing confusion over the presence of the white women and their role in the action and bringing content knowledge and relevant vocabulary to bear on this analysis, especially concerning poll taxes, literacy tests, 15th amendment, women's fight for the vote, boycotts, and protests.

The comparison group was not able to interpret the photo in any accurate context. Students in the group suggested that the photo was of "when black people and white people got together", "they're trying to figure out where they live", and people trying to make it so that there "could be some – little bit of slavery and maybe they want slavery because they don't want to do their work and stuff." One student thought the sign in the photograph said, "End the slavery test" (instead of "End the literacy test") and when the interviewer helped them read it and asked what "End the literacy test" might mean, several students thought this probably had something to do with littering.

5.3 Other Outcomes

5.3.1 Dissemination

One of the *BHH* project goals was to disseminate the *BHH* curriculum and pedagogical methods to a wider audience. This goal has been fulfilled in several ways. First, *BHH* has a webpage with the entire elementary curriculum available to anyone interested in accessing the materials. The website is found at www.bringinghistoryhome.org. In addition, the funding of the

BHHIII – Grant Wood History Institute has extended the use of the *BHH* pedagogical methods to middle school and high school history instruction and descriptions of the paradigm are available on the website.

Second, the curriculum has been adopted or in the process of adoption for use in schools in several states. The curriculum is the official elementary social studies curriculum in the *BHHII* districts and in addition to the project schools, in Iowa, the BCLUW Community School District in Conrad and the Prairie Ridge Elementary School in Cedar Rapids have also adopted the curriculum. The five pedagogical processes of the *BHH* curriculum are in use in connection with a TAH grant in Plymouth, MI. During 2006, the curriculum was introduced and is under consideration for use in Galesburg, IL and the Suburban School District Consortium in St. Louis, MO. Finally, Fargo, ND Community School District has designed a K-5 history program that makes use of the BHH pedogogies.

The *BHH* curriculum has also been disseminated through conference presentations and journal articles. Table 24 lists the conference presentations given by the project director and teacher participants during the grant years, from 2003 to the present. Table 25 shows the journal articles during that time. The Center for Evaluation and Assessment is working with the project director on additional research publications to be submitted in the next year.

Table 24. Bringing History Home Conference Presentations

Year	Title of Presentation	Conference
2003	*Community Glue: Extracurricular Social Studies Programs*	National Council for the Social Studies (NCSS) Annual Conference
	High Expectations and Deep Foundations: Real Teachers, Real Classrooms, Real History in K-3	National Council for History Education (NCHE) Annual Conference
2004	*Using Technology to Teach History* (Panelist)	Teaching American History (TAH) Directors' Meeting
	Westward Expansion: A 3rd Grade Unit Designed by 3rd Grade Teachers	Iowa and Nebraska CSS Annual Meeting
	Bringing History Home (Featured presentation)	Iowa Educational Research and Evaluation Association (IEREA), Annual Meeting
2005	*Designing Assessments of Student Learning in History* and *What can K-3 Children Learn about History?*	NCSS: Great Lakes Regional Conference
	Linking History Content Institutes to Classroom Learning (Panelist)	TAH Annual Director's Meeting
	Start with the Text, End with Historical Thinking and *Assessing Student Learning in History*	Iowa Council for the Social Studies (ICSS), Annual Conference
	Designing an Elementary Teaching American History Project	NCSS, Annual Conference

2006	*The Bringing History Home Project*	NCHE Annual Conference
	TAH Project Dissemination (Presenter) and *Regional Directors' Meeting* (Session leader)	TAH Annual Directors' Meeting
	Historical Inquiry Activities	ICSS Annual Conference
2007	*When Historians and Teachers Collaborate: Lessons from Teaching American History Grant Partnerships* (panelist), *North Central TAH Grant Directors' and Teachers' Meeting* (Session Leader), and *Elementary History Education* (Roundtable Discussion Leader)	Organization of American Historians (OAH)

Table 25. Bringing History Home Journal Articles

Year	Title of Article	Journal
2004	*Teaching History K-6*	Iowa Council for the Social Studies Journal, Vol. 17, No. 1, 17-21.
2005	*Start with the Text, End with Historical Thinking*	Iowa Council for the Social Studies Journal, Vol. 18, No. 1, 11-17.

5.3.2 Mentor Involvement

Another objective of the *BHHII* project was to foster mentors who could help with project dissemination by serving as grade level experts at professional development workshops for new schools that might become involved in teaching the curriculum, and who would serve as leaders for grade level teaching peers. During the course of the *BHH* project, the project director recruited mentors to work with teachers as part of the *BHHII* project and over the course of the *BHHII* project, additional teachers from the new schools became enthusiastic participants in mentoring. In all, 15 teachers have led sessions at professional development workshops. Several of these teachers have received their Masters Degrees during the last few years and are now eligible to lead *BHH* workshops for continuing education credit through the ten Iowa Area Education Agencies. Six of the teachers have participated in conference presentations at the Iowa or National Council for the Social Studies meetings.

Many of the mentors have also taken the initiative to write additional history curricular units for the elementary grades, and several teachers responded to the project director's offer to share their adaptations of the curriculum. (Many teacher-written adaptations are on the project website.) Other teachers, while not traveling beyond their own school as mentors, have chosen to be leaders for their grade level peers in their home schools, gathering materials, sharing websites, and coordinating their enthusiastic efforts in teaching history.

Several teachers in the project reported privately (in conversations with evaluation staff members) and publicly (at professional development events and conferences), that participating in the *BHH* project and becoming excited about teaching history had rejuvenated them as teachers. One teacher, Kimberly Heckart, who carried the curriculum with her to a new district when she moved to a new district, has collected accolades as a history teacher since becoming part of the project. She was named the 2006 Iowa Council for the Social Studies Elementary Teacher of the Year, the 2007 National Council for the Social Studies Elementary Teacher of the Year, and the 2007 Gilder Lehrman Institute of American History Iowa History Teacher of the Year. In accepting the Gilder Lehrman award, she used and cited the format of the Martin Luther King, Jr. "*I have a dream"* speech to tell her students and fellow teachers (excerpted from her talk):

I have a dream today! I have a dream that children are inspired with a love of history learning in every grade. Thank you to the students at Prairie Ridge for showing your interest in history and your inquiring minds that question and make you want to learn more. I am proud to be one of the teachers of such smart students!

I have a dream today! I have a dream that all children in our nation will have the opportunity to learn about history by teachers who use an inquiry-based learning. Teachers who use strategies to challenge kids to think critically: through photo analysis, document analysis, time lining, mapping, and synthesizing projects. Thank you to the teachers of Prairie Ridge for leading in this challenge. I am proud to be teaching here with each of you.

I have a dream today! I have a dream that everyone in the world could have a mentor and friend as I have found in Elise Fillpot, who is the project director for the Bringing History Home Teaching American History Grant. Without Elise and Bringing History Home, I would not be the history teacher that I am today.

6. Conclusions

The goals of the BHH II Project were to:

1. Provide professional development in source-based historical inquiry using the twelve curricular units designed as part of the BHH project
2. Implement the BHH curriculum in K-5 classrooms in three new districts in the state of Iowa.
3. Disseminate the BHH curriculum through the creation of a network of K-5 history advocates in Iowa and contribute to national awareness of high quality elementary history education.

The BHH II project accomplished all of these goals and has contributed to the body of research concerning elementary children's ability to learn historical content and practice historical thinking skills. Summer professional development workshops and ongoing professional development support provided during the academic years of the project gave approximately 123 K-5 teachers the confidence and knowledge to teach twelve new history units to approximately 3000 students each year. The BHH curricula and pedagogical approach have been adopted by all these districts, and several additional schools and districts have adopted at least part of the curricula or approach.

Teacher surveys, interviews, and classroom observations documented implementation of the history units in the K-5 classrooms. BHH II teachers believed that their students had learned new historical content and acquired skills to facilitate future history learning. In the BHH II schools, there was more history instruction than there had been prior to the project, and there was more history instruction in the four treatment schools than there was in two other comparison Iowa school districts. Participating teachers became convinced that their students were able not only to learn historical content, but with practice in guided-inquiry, could also acquire historical thinking skills that would help them understand history. Teacher responses indicated that some BHH curricular units were preferred over others, either because of their own interest (or lack thereof) in the content area, or because of the teachers' perceptions of the appropriateness, difficulty, or inherent appeal of the materials for use with their students. The BHH curriculum continues to evolve and be improved as more teachers provide feedback on how it works in the classroom.

Evaluation activities documented student outcomes (for 3rd through 5th grade students) through two types of assessments completed by students; written narrative assessments using key words from each curricular unit, and photo analysis assessments using photographs from historical eras addressed in each curricular unit. Evaluators observed classrooms in all six grades and

conducted group interviews with small samples of students in grades 3 through 5. As a result of exposure to the BHH curricular units, students in BHH II treatment schools showed strong improvement in their ability to construct historical narratives and to use photographs to learn and display their command of historical content knowledge and skills. In addition to demonstrating increased knowledge and ability to use historical thinking skills over time, students in treatment schools outperformed their comparison group peers on all measures. Students' performance gains were not even across all measures, in several cases mirroring the teachers' perceptions of which units worked better or less well for students. Curriculum revisions occurred throughout the course of the grant and continue as the project continues to be disseminated.

The BHH II intervention was successful in providing professional development in teaching history to elementary teachers, implementing history instruction in elementary classrooms, disseminating the BHH history curricula and pedagogical approach, and improving the historical content knowledge and ability to think historically of elementary school children.

7. References

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38 (4), 915-945.*

Linn, R.L. & Miller, M.D. (2005). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Prentice Hall.

Mehrens,W.A. & Lehman, I.J. (1991). *Measurement and evaluation in education and psychology,* Orlando, FL: Harcourt Brace Jovanovich.