



RESEARCH REPORT

Effectiveness of TCI's *History Alive!*
for Eighth Graders:
A Report of a Randomized Experiment in
Alum Rock Union Elementary School District

Jessica Villaruz Cabalo
Empirical Education Inc.
and University of California, Santa Barbara

Denis Newman
Empirical Education Inc.

Andrew Jaciw
Stanford University

March 7, 2006

Empirical Education Inc.
www.empiricaleducation.com
425 Sherman Avenue, Suite 210
Palo Alto, CA 94306
(650) 328-1734

Acknowledgements

We are grateful to the people in the Alum Rock and the Berryessa school districts for their interest and assistance in conducting this research. Teachers' Curriculum Institute graciously provided all the classroom materials and training at no charge to the school districts. The research was funded by a grant (#R305E040031) to Empirical Education Inc. from the US Department of Education. The purpose of this grant is to improve our ability to run small scale experiments to assist local decision-makers. The US ED is not responsible for the content of this report.

About Empirical Education Inc.

Empirical Education Inc. was founded to help K–12 school districts, publishers, and the educational R&D community assess new or proposed instructional programs through scientifically based pilot implementations. The company draws on the expertise of world-class researchers and methodologists assuring that the research is objective and takes advantage of current best practice in rigorous experimental design and statistical analysis. The company's findings let educators quantify the value of programs and help them partner with providers to implement those most effective for their students.

© 2006 by Empirical Education Inc. All rights reserved.

Reference this report: Cabalo, J.V., Newman, D., & Jaciw, A. (2006, March). Effectiveness of TCI's History Alive! for eighth graders: A report of a randomized experiment in Alum Rock Union Elementary School District. Palo Alto, CA: Empirical Education Inc. Retrievable from https://www.empiricaleducation.com/past_research/

Executive Summary

Alum Rock Elementary Union School District had an interest in a middle school program being introduced by Teachers' Curriculum Institute (TCI) called *History Alive!* Supplementary materials from TCI were already in use in many of the classrooms and several of the district administrators were interested in whether the full middle school program recently introduced would be successful in their schools. *History Alive!* takes an activity-oriented approach quite distinct from a conventional textbook program. It was important to know whether this kind of program would be as effective as the current approach in achieving student success on the California Standards Test (CST) for history.

History Alive! *History Alive!* The United States program utilizes the TCI approach based on multiple intelligences, cooperative interaction and spiral curriculum. In addition to the student textbook that the regular history program utilizes, the *History Alive!* program is activity-oriented and incorporates use of additional materials for teachers and students. Professional development is also an important part of the program.

Setting. Alum Rock Elementary Union School District is a large (13,600 student) district with 20 elementary schools (K-5) and seven middle schools (1 K-8 and 6 6-8). The student population is approximately 76.3% Hispanic, 10.4% Asian, 3.5% White, and 3% African American. 80% of the student population was classified as economically disadvantaged and 59% of the student population was designated as English learners.

The district had previously adopted a textbook program as its regular history program for 8th graders prior to the availability of the TCI full-year *History Alive!* program. In seeking to adopt new history texts, the district recognized the need to test the effectiveness of a history program that employed more interactive techniques and materials.

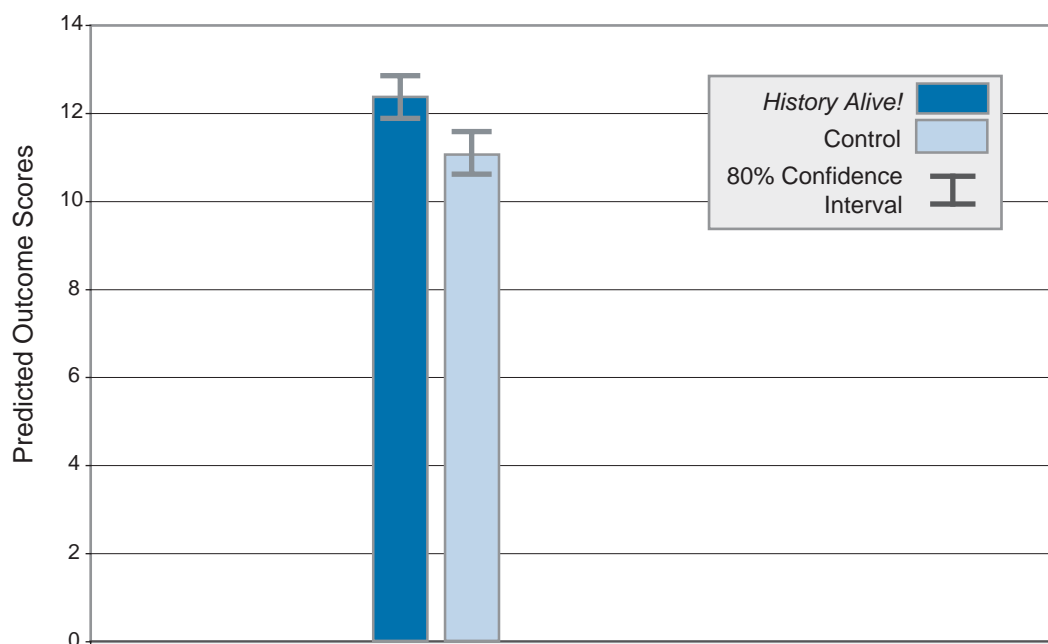
Research design. The research is a comparison of outcomes for groups of students taught using the *History Alive!* program and similar students taught using the districts' regular textbook-based history program. To help assure that student demographics were distributed between the both groups, pairs were formed within schools. Where more than two teachers came from the same school, teaching experience was used and as a consequence two teachers with substantial TCI experience were paired. Between each pair of teachers, we used a coin toss to randomly assign one of the teacher volunteers to the *History Alive!* group and the other to the control group. The primary outcome measure was the California Standards Test (CST) for History-Social Science (History) results from 2005.

Participants. After two teachers were removed for reasons unrelated to the treatment, there were four *History Alive!* teachers and five control teachers in the study. Some of the teachers had prior training and exposure to TCI's instructional approach and program materials. We measured this "contamination" through surveys.

Statistical analysis. We used a mixed model statistical analysis that involved two levels-- students and classes. We conducted three analyses. First we looked at the outcome measure, CST History, controlling for prior CST scores in English Language Arts (ELA). Second, we conducted the same analysis except controlling for English proficiency. Finally, we examined ELA as an outcome controlling for prior ELA score.

Results. The three analyses yielded a common pattern. In each case there was essentially no impact for the average student and in no case did we see a substantial negative impact for higher achieving

students. But for the lower achieving students, there was a positive impact for History and ELA outcomes. We found a similar result but not as strong when looking at the difference between English learners and others. The following bar graph shows the statistical model's prediction for the student at the median of the bottom quartile of ELA achievement.



Median Student of the Bottom Quartile of the Pretest

Conclusion. Our randomized experiment in this district provides evidence of a positive impact for their lower scoring students working with TCI's *History Alive!* in comparison to what can be expected with the conventional textbook programs. Although there was no evidence of an advantage for the average student for which both programs gave similar results, we consistently found an interaction between the condition and the pretest score.

The limitations of this study must be considered in the interpretation. The prior use of the TCI materials by many of the control teachers would be expected to lower the contrast between the *History Alive!* and control groups and possibly make the apparent impact smaller than it might have been. Importantly, with only nine teachers altogether, there is a danger of bias introduced by chance. With respect to the possibility that the condition by ELA achievement interaction was due to an imbalanced distribution of teachers, we showed that it is unlikely to have occurred by chance. Our result for the English Language Arts outcome is intriguing and warrants further investigation. We see the result as providing support for the conclusion that *History Alive!* is differentially effective for the lower scoring students.

The important finding was that *History Alive!* differentially benefits the students with lower ELA scores and possibly those who are learning English. In districts with large numbers of such students, this program may have the effect of reducing the achievement gap. Considered as a local pilot in Alum Rock, the study adds to the information available on which to base their adoption decisions.

Effectiveness of TCI's *History Alive!*
for Eighth Graders:
A Report of a Randomized Experiment in
Alum Rock Union Elementary School District

Table of Contents

Acknowledgements	i
Executive Summary	ii
Introduction	1
Methods	1
RESEARCH DESIGN	1
MATERIALS	2
SITE DESCRIPTION	2
SAMPLE AND RANDOMIZATION	2
DATA COLLECTION	3
Test Scores	3
Surveys	4
STATISTICAL ANALYSIS	4
Results	4
FORMATION OF THE EXPERIMENTAL GROUPS	4
Table 1: Distribution of <i>History Alive!</i> group broken down by schools, teachers and counts of students	5
Table 2: Distribution of control group broken down by schools, teachers and counts of students	6
Table 3: Independent t-test of the difference between <i>History Alive!</i> and control groups for the English Language Arts pretest	6
Table 4: Chi square table of the distribution of students with missing ELA pretest data between <i>History Alive!</i> and control groups	7
Table 5: Chi square table of the differences between English learners and non-English learners in <i>History Alive!</i> and control groups	7
ATTRITION	8
PROGRAM IMPLEMENTATION	9
STATISTICAL MODELS FOR THE OUTCOME MEASURES	9
RESULTS FOR HISTORY CONTROLLING FOR ELA	9
Table 6: Multi-level mixed model for History--results controlling for ELA pretest	10
Figure 1: History outcome --scatterplot showing <i>History Alive!</i> and control students with lines showing the predicted values based on pretest score	10
Figure 2: History outcome—difference between <i>History Alive!</i> and control with the predicted values for the median student at each quartile of the pretest	11
Figure 3: History outcome--bar graph showing the difference between <i>History Alive!</i> and control groups for the median student of the bottom quartile	12
RESULTS FOR HISTORY CONTROLLING FOR ENGLISH LEARNER STATUS	13
Table 7: Multi-level mixed model for History—results controlling for English learner status	13
Figure 4: History outcome--bar graph showing the difference between English learners and non-English learners in both <i>History Alive!</i> and control groups	14
RESULTS FOR ENGLISH LANGUAGE ARTS	15
Table 8: Multi-level mixed model for ELA—results controlling for ELA pretest	14
Figure 5: English Language Arts outcome--scatterplot showing <i>History Alive!</i> and control group students with lines showing the predicted values based on pretest score	15

Figure 6: English Language Arts outcome—difference between *History Alive!* and control with the predicted values for the median student at each quartile of the pretest. 16

Figure 7: English Language Arts outcome--bar graph showing the difference between *History Alive!* and control group for the median student of the bottom quartile. 16

Discussion	17
References	18
Appendix A	19
Appendix B	22

Introduction

This research on the effectiveness of one of the programs currently up for adoption in California grew out of discussions among several school districts, the Santa Clara County Office of Education, Teachers' Curriculum Institute (TCI) and Empirical Education. The US Department of Education's research funds supported Empirical Education's efforts in the research. Alum Rock Elementary Union School District had an interest in a middle school program being introduced by TCI called *History Alive!* Supplementary materials from TCI were already in use in many of the classrooms and several of the district administrators were interested in whether the full middle school program recently introduced would be successful in their schools. Measures of effectiveness do not play a role in the state adoption process, which focuses on the program's meeting a stringent set of content and format standards. *History Alive!* met these standards but takes an activity-oriented approach quite distinct from a conventional textbook program. It was important to know whether this kind of program would be as effective as the current approach in achieving student success on the California Standards Test (CST) for History. A measure of the impact of the program could provide useful evidence to support district decisions about which history program to adopt.

We conducted an experiment in 28 eighth grade classes in Alum Rock. We randomly assigned eighth grade history teacher volunteers to either use the *History Alive!* program (the pilot group) or to continue using the currently adopted textbook program (the control group). The *History Alive!* teachers used *History Alive!* for eight months from the beginning of the 2004-2005 school year until the CST in History was administered in April 2005.

The question we addressed specifically is whether students in classes that use *History Alive!* materials will get higher scores on the history assessment than they would if they had been in a control classroom. We were also interested in the possibility that *History Alive!* could compensate for lower student reading ability since it used classroom activities and many supplementary materials to help students reflect on the content. So we also started with the question as to whether there may be a different impact for students at different levels of ELA achievement. The district had a large English learner population as well as students with relatively low scores in English Language Arts (ELA), so we also wanted to know whether there was any differential effect of the program depending on the student's incoming command of English.

Our experimental design reflects the requirements of the No Child Left Behind Act, which directs schools to consult reports of rigorous research to guide their adoptions of instructional programs. This study was designed to provide useful information to support a local decision in Alum Rock but not to generate broadly generalizable results. The results should not be considered to apply to school districts with practices and populations different from those found in Alum Rock. In addition, because of the small number of teachers involved, the local decision-makers must consider carefully whether those teachers are a good representation of their staff as a whole.

Methods

Research Design

The study is a comparison of outcomes for groups of students taught using TCI's *History Alive!* program (the *History Alive!* group) and students taught using the regular textbook-based history program (the control group). The design uses randomization process. We randomly assigned teachers to one condition

or the other. Each teacher had one or more 8th grade history classes, all of which were designated to be in either the *History Alive!* or control group depending on the teacher assignment. The primary outcome measure was the student score on the California Standards Test (CST). We controlled for initial characteristics of the student primarily through their previous year's score on the English Language Arts part of the CST and whether they were English learners.

The experiment started at the beginning of the 2004-2005 school year. We based our analysis on nine teachers, 28 classes and 820 students. The number of teachers that were willing to try out the new program was smaller than we had initially planned. Nevertheless, we proceeded with the experiment knowing that the program would have to have a relatively large impact to be detectable. We also understood that with a small number of teachers we have to be cautious that teacher differences if unevenly distributed between the *History Alive!* and control groups, can bias the result.

Materials

As described by the Teachers' Curriculum Institute (2006) *History Alive!*: The United States program is a full year program based on an approach to instruction. According to the publisher, the program is characterized by theory-based active instruction, standards-based content, multiple intelligence teaching strategies, considerate text, graphically organized reading notes, preview and summary assignments for each lesson, and multiple intelligence assessments. The TCI approach incorporates standards-based instruction, while promoting innovative, effective instruction that excites students about social studies (TCI, 2006). The program materials include lesson guides, interactive student notebooks, and supplemental materials such as maps, overheads, placards, compact disc, desk map, interactive timeline and enrichment resources and projects.

Their current textbook-based program also provided interactive and interdisciplinary activities to foster student engagement. The primary differences between the two programs was that the TCI's *History Alive!* program supplies additional classroom materials and has a greater focus on an approach specifically intended to move away from lectures, recitation and seatwork.

History Alive! teachers received their complete set of classroom materials during the initial meeting. These teachers also attended three days of professional development led by a TCI consultant. Beyond the initial training, teachers were free to make use of the materials as best suited the needs of their classroom and students.

Site Description

Alum Rock Elementary Union School District is part of the Santa Clara County located in the city of San Jose, approximately 48 miles southeast of San Francisco, CA. It is a relatively large (approximately 13,604 student) district with 20 elementary schools (K-5) and seven middle schools (1 K-8 and 6 6-8). The student population is approximately 76.3% Hispanic, 10.4% Asian, 3.5% White, and 3% African American. Eighty percent of the student population was classified as economically disadvantaged and 59% of the student population was designated as English learners. The district had previously adopted a textbook as its regular history program for 8th graders prior to the availability of the full-year *History Alive!* The United States program.

Sample and Randomization

A member of the TCI sales force initially introduced the researchers to several districts in Santa Clara County interested in the TCI *History Alive!* program and of those, two were interested in conducting a

structured pilot with some of their classes. We decided to conduct the experiment in grade 8 since that is the year that the state test for history (in California this is called “History–Social Science”) is given. Researchers met with these district staff to explain the details of the study. Principals invited the interested 8th grade teachers to an after school meeting at which researchers introduced the study on the *History Alive!* program and held a discussion about the research procedures.

Eleven teachers attended the initial meeting for the experiment in Alum Rock in August 2004. Two additional participants from neighboring Berryessa were recruited the week after the initial meeting. After a question and answer period at the Alum Rock meeting, teacher volunteers engaged in a discussion of the important factors that they believe will have an impact on the results in their district. The high percentage of English learners was raised as an important issue in this district. Teachers also pointed out that some of them already had prior training in and used some TCI materials. Because of this, some teachers suggested that those with such experience should be specifically assigned to the *History Alive!* group. After detailed explanation of the design and premise of a randomized study, randomization proceeded.

To form similar groups, we paired the teachers and we used a coin toss to randomly assign one of the teacher volunteers to the *History Alive!* group and the other to the control group. The pairs were formed within schools. Where there were more than two teachers in the school, pairs were formed based on teaching experience. Where there were an uneven number, the coin was tossed separately for the final member.

Of the 13 teachers originally signed up, only one teacher, assigned to the *History Alive!* group, was dropped from the study because he switched grade levels. The two teachers from neighboring Berryessa were also dropped, as explained in the results section, as we focused this report on the results for Alum Rock only. Finally, one *History Alive!* teacher was dropped when we discovered that her classes were for very high achieving students and not representative of the larger Alum Rock population. The final count was four *History Alive!* teachers and five control teachers.

Data Collection

Test Scores

Alum Rock USD provided us with the test scores and demographic data. We used the California State Test (CST) scores for History and English Language Arts (ELA). The CST is a criterion-referenced test and results are based on how well students achieve identified state-adopted content standards. The CST in history covers the standards for grades 6, 7, and 8 and is only administered to students in 8th grade. Because of this, we did not have any pretest score in history. However, we did have pretest measures for ELA which we believed would be highly correlated with history achievement and therefore useful as a covariate to help control for the student’s incoming academic preparation.

For pretest measures, we used the CST scores in ELA from April 2004 as well as English proficiency categories from 2004. The outcome measure for history was based on two of the topic areas in the California Standards Test for History—Social Science. This test given in eighth grade examines all the material covered in middle school (grades 6 through 8). Of the five topics covered, two are addressed in eighth grade. Student results for each topic are reported separately in terms of a raw score (and number of items) for each of the topics. Since our intervention occurred only in eighth grade, we selected the two relevant topics: U.S. Constitution and Early Republic and the Civil War and the Aftermath. The first topic was tested with 22 items and the second with 13 items. For each student we added the number correct for each topic generating a raw score out of 35 as our outcome measure.

Since we found that student scores on the two topics were correlated at .68 and that using a scale that weighted the topics equally gave the substantially the same result, we followed the simple procedure of adding the raw scores. We can consider the test result to be based on a sample of 35 test items that address eighth grade standards and a reasonable measure of achievement for the year.

The English proficiency scores were broken down into four categories: native English, fluent English proficient (family speaks another language but the student came to school fluent in English), limited English (or English learner) and re-designated as proficient (students who previously had been coded as English learners). We collapsed three of the categories to get a dichotomous variable of English learner and non-English learner.

Surveys

Through three web-based surveys given to participating teachers in both groups, researchers tracked the usage of the TCI *History Alive!* materials, student engagement and class interactions with the materials. See Appendix A and B for sample surveys. Surveys began mid-year after teachers had already been implementing a history program for 4 months.

Statistical Analysis

Our primary outcome measure was the CST in History. The basic question for the statistical analysis was whether students in the *History Alive!* classrooms had higher history scores than those in the control classrooms. Recognizing that whether the teacher was piloting the new program or not is not the only factor influencing the results, we developed statistical models that took into account the student's pretest score as well as whether or not they were English learners. An analysis of covariance allows us to look at these variables (covariates) simultaneously and to identify how the variables individually and in combination impact the outcome. The statistical models were multi-level because they accounted for the clustering of students in classes, which provides a more accurate, and often more conservative, assessment of the confidence we should have in the findings. We explicitly modeled the levels for which there was a significant amount of variance to be explained. We decided which covariates to include based on our prior expectations about factors that should make a difference. Beyond these, we construct exploratory models to better understand unexpected results. We use SAS PROC MIXED (from SAS Institute Inc.) as the primary tool for this work.

Results

Formation of the Experimental Groups

The randomizing process helps ensure that our estimates from the experiment are unbiased, but does not guarantee that the groups will be perfectly matched on all characteristics. It is important to inspect the two groups to see if any significant differences occurred that would have to be controlled statistically.

The following tables address the nature of the groups as initially formed. Table 1 shows the distribution of students between control and *History Alive!* conditions, and the distribution of classrooms in schools. The number of classes was not a criterion for teacher pairings during the randomization. As a result, the distribution of classes and students was not balanced. Teachers who had more classes landed in the control group.

Table 1: Distribution of Schools, Teachers and Students

School ID#	Teacher ID#	Class ID#	Number of Students
504	531	1221	32
506	528	1227	28
		1233	36
		1234	35
507	526	1235	32
		1236	32
		1237	32
		1238	31
		1240	31
508	533	1254	30
		1255	31
	Total Teachers = 4	Total Classes = 11	Total Students = 350

Table 2: Distribution of control group broken down by schools, teachers and counts of students

School ID#	Teacher ID#	Class ID#	Number of Students
504	532	1241	31
		1242	30
		1243	17
		1244	31
505	525	1245	31
		1213	31
		1214	33
		1215	27
		1216	31
		1217	23
506	529	1222	32
508	534	1207	33
		1208	22
		1209	28
	527	1210	35
		1211	15
		1212	20
Total Teachers = 5		Total Classes = 17	Total Students = 470

We can also compare the *History Alive!* and control students on variables that may be relevant to the analysis as shown in the following tables. Table 3 is an independent samples *t* test that shows there is an initial difference between the *History Alive!* and control groups on the ELA pretest. The measure of “effect size” is an indication that the difference, compared to the overall amount of variation among the students, could impact the results. The effect size expresses this discrepancy, which is due to chance, in standard deviation units. Since the control group was starting with higher scores, a statistical adjustment will have to be made.

Table 3: Independent *t* test of the difference between *History Alive!* and control groups for the English Language Arts pretest

Descriptive statistics: CST English Language Arts pretest	Raw Group Means	Standard Deviation	Number of Students	Standard Error	Effect Size
<i>History Alive!</i>	313.48	42.596	277	2.559	-0.258
Control	325.11	47.429	369	2.469	
t-test for difference between independent means	Difference		DF	<i>t</i> value	<i>p</i> value
Condition (<i>History Alive!</i> – control)	-11.63		644	3.22	0.0013

Note: 174 students were missing the ELA pretest

It is important to note that of the 820 students, 174, or 21.2% were missing data for this pretest. This will narrow the cases available for analysis using this score as a covariate. We also conducted a statistical test to see if these missing students were distributed differently between the two conditions. Table 4 shows the results of this Chi Square test. The high p value indicates that there is a high probability the small difference in proportion is a result of chance. Similar tests comparing English learners and others in their distributions of missing pretest scores also showed no difference.

Table 4: Chi square table of the distribution of students with missing ELA pretest data between *History Alive!* and control groups

Condition	Has a pretest	Pretest missing	Totals
<i>History Alive!</i>	277	73	350
Control	369	101	470
Totals	646	174	820
Chi-square statistics	DF	value	p value
	3	.02	0.8945

We were also concerned whether English learners were evenly distributed between the two conditions. Since over half of the student population in the study were English Learners, English was expected to be an important factor in the analysis. Table 5 reports a Chi Square test of the distribution of English learners vs. non-English learners in the *History Alive!* and control groups.

Table 5: Chi square table of the differences between English learners and non-English learners in *History Alive!* and control groups

Condition	English Learner	Non-English Learner	Total
<i>History Alive!</i>	222	128	350
Control	229	241	470
Total	451	369	820
Statistic	DF	value	p value
Chi-square	1	17.527	<0.0001

Note: 0 students were missing English proficiency information.

This table is consistent with the information in Table 3, which indicated the control group on average scored higher on the ELA pretest. Here the *History Alive!* group had 63.4% English learners while the control group had only 48.7%. It will be important to explore the treatment effect controlling for the potential confounding effect of English learner status.

Attrition

Of the 11 Alum Rock teachers who were randomized, one *History Alive!* group teacher was dropped from the study due to reassignment unrelated to the experiment. We dropped a second *History Alive!* teacher after pretests and survey comments indicated that her classes were specifically for very high achieving students. There were no corresponding classes in the control group. Classes for these teachers are not shown in Tables 1 and 2. Of the 820 students in the initial sample, all but 3.2% also took the history posttest, a low rate of attrition.

Program Implementation

At the beginning of the school year, the *History Alive!* teachers attended 3 intensive days of professional development workshops focusing on teaching strategies and curriculum development. These workshops were conducted by a TCI consultant who trained the teachers on the TCI approach to improving student achievement. These topics included: effective ways to increase engagement and interaction, work with multiple intelligences and build students' content reading skills. Other topics covered included creating a cooperative classroom environment, developing lessons and using the *History Alive!* materials to better meet the needs of state and district standards.

History Alive! teachers utilized this knowledge to implement the *History Alive!* program in their classrooms. Teachers reported general ease of use of the program as well as positive interactions with the student and teacher materials. One teacher, however, reported difficulty in implementing particular activities with large groups of students. The teacher of the high achieving students who was dropped reported that her students took to the program readily.

Control teachers relied on the use of their textbook as well as supplemental material that they sought out themselves through the internet and/or other non-TCI resources.

Among the survey respondents, mostly all *History Alive!* teachers reported using the *History Alive!* materials 75-100% of the time during instructional time in history. One exception to this is a reported use of 40% by one teacher. During this time, *History Alive!* teachers consistently used the student book and notebook, transparencies, placards, CD, desk map, activities, lesson guide and assessments. Overall, *History Alive!* teachers reported positive responses to the *History Alive!* program.

Control teachers reported using some of the TCI materials (obtained prior to the study) 0 to 10% of their instructional time in history. When they did use these materials, control teachers used them as supplementary to their regular history textbook. Materials used included the student notebook, transparencies, CD and activities.

On a Likert scale of 1 to 5 (1 indicating 'not engaged' and 5 indicating 'fully engaged') teachers in the *History Alive!* classrooms reported a 3.99 overall average of student engagement. In comparison, teachers in the control classrooms reported a 3.2 overall average of student engagement.

Statistical Models for the Outcome Measures

We develop statistical models that take into consideration the clustering of students into classes, teachers, schools, and initially districts. We explicitly modeled levels for which there was a significant amount of variance to be explained. In our initial analysis, we were working with the Alum Rock data as well as the classes for two teachers from neighboring Berryessa. These teachers had joined the study a week later than those from Alum Rock and were randomized as a pair. Initially we believed that adding

additional units to the experiment would help by increasing the statistical power and, given the districts were adjacent, the two districts could be considered as one. However, in the initial look we found that there was a difference between the two districts in average outcomes. On further inspection, we found that the classes from Berryessa (like the district as a whole) contained a different demographic, a much higher portion of Asian students among their English learners. Since our initial goal was to provide evidence that would help the Alum Rock decision-makers, including the Berryessa students complicated the analysis and brought in factors that were not representative of Alum Rock. For this reason, we chose to drop those two teachers and focus the analysis only on the district from which most of the classes were drawn.

Our main outcome variable was the score in History, which consisted of the combined raw scores of the two relevant sub-topics for eighth grade. Our primary covariate that allows us to control the students' incoming achievement was their ELA score. Since we were also interested in whether *History Alive!* might compensate for lower reading ability, we wanted to test if there was a different impact for students at different levels of ELA achievement. The same thinking carried over to English proficiency. Since English and proficiency and ELA were related, we chose to construct two separate models, one controlling for ELA and the other for English learner status, in both cases considering whether there would be a differential impact.

We also had the English Language Arts outcome measure for all of the students. While it was not initially evident that the history program would have an impact on ELA, which is taught by different teachers, we chose to model that outcome as well on an exploratory basis.

While we inspect multiple models involving combinations of the variables of initial interest, we report only the model that we believe provides the best and most parsimonious account of the results.

Results for History Controlling for ELA

We address the results for the CST History scores using two different covariates: the CST scores in English Language Arts (ELA) from 2004 as well as English proficiency level.

We first looked at the results for History using ELA scores as a covariate. We used the ELA scores because there were no prior history scores in 7th grade and we believed it to be a good predictor of the History outcome.

Table 6 displays both the descriptive statistics including the raw means for the two conditions and the analysis of these results using the statistical model that includes the pretest. The bottom segment of the table presents technical information on how the model takes the clustering of students into classes into account. Of interest here are the lines for condition and for condition by pretest interaction. The condition accounts for about 0.13 of a point on the raw history scale and the p value is very high indicating that it is reasonably likely that there is no difference. The pretest score was centered on the average so this difference applies only to the average student. But we have students at many different levels in the English Language Arts achievement and the much lower p value for the condition by pretest interaction indicates that there is likely to be a difference that depends on the incoming ELA score. A p value of .01 indicates that there is about a 1 in 100 chance that an interaction this large would occur just by chance in our small sample if there was not actually a difference to be measured.

Table 6: Multi-level mixed model for History--results controlling for ELA pretest

Descriptive statistics: CST History outcomes	Raw Group Means	Standard Deviation	Number of Students	Number of Classes	
<i>History Alive!</i>	326.41	41.901	343	11	
Control	328.3	48.853	454	17	
Mixed model: Fixed factors related to CST History outcomes	Estimate of Coefficient	Standard Error	DF	t value	p value
Intercept	16.230	0.467	26	34.77	<.0001
Pretest score (centered at the mean)	0.101	0.005	604	20.64	<.0001
Condition (<i>History Alive!</i> = 1; control = 0)	0.133	0.737	604	0.18	0.857
Condition by pretest interaction	-0.022	0.009	604	-2.54	0.011
Mixed model: Technical details for random components	Estimate of Variance Component	Standard Error		z value	p value
Class mean achievement	2.872	1.0087		2.85	0.002
Within class variation	16.379	0.942		17.38	<.0001

Note: this model is based on 634 cases. Missing cases include 183 students without either pretests or posttests. Another 3 cases were removed as outliers or influential points.

These results warrant a closer look at the nature of this interaction. As a visual representation of this result, we present a scatterplot in Figure 1 that shows where the students fell in terms of their starting

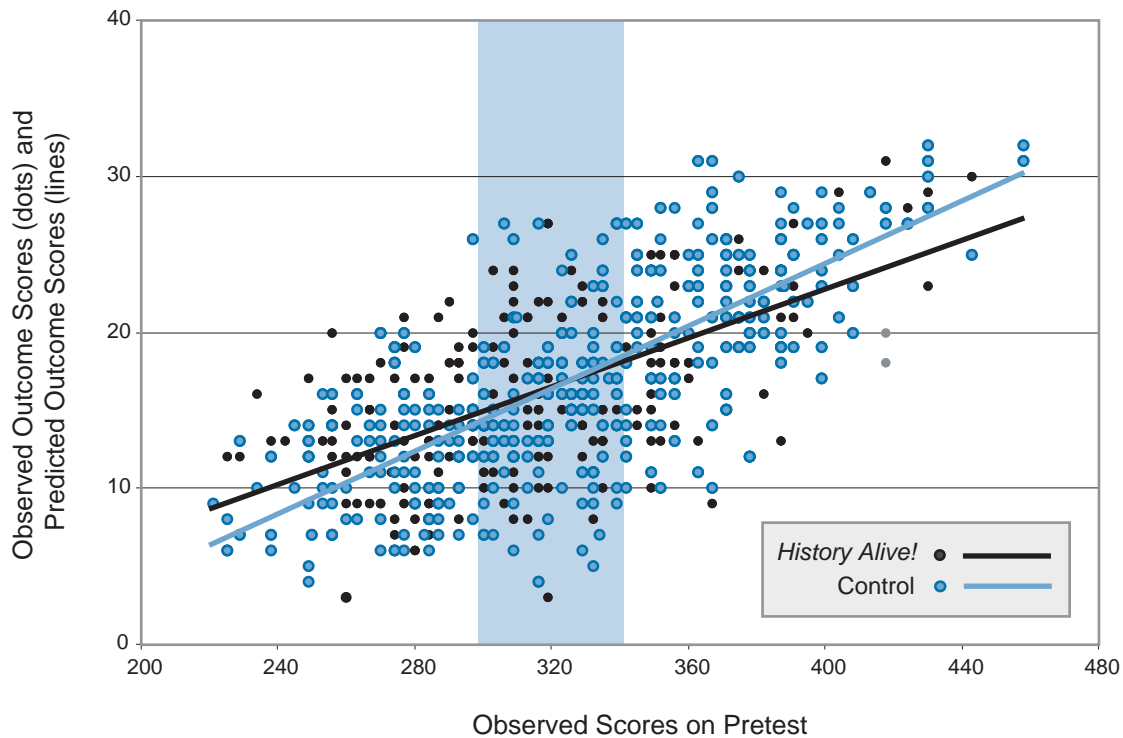


Figure 1: History outcome --scatterplot showing History Alive! and control students with lines showing the predicted values based on pretest score

point (horizontal x-axis) and their CST history score (vertical y-axis). This graph illustrates the differences in growth among the students. The two lines represent the *History Alive!* (dark line) and control (light line) groups in terms of what the statistical model predicts a student's outcome score will be, given where he or she started on the pretest scale.

For reference, we provide information on the ELA proficiency levels set by California. For 7th graders, “far below basic” are scores below 262. “Below basic” ranges from 263 to 299. “Basic” ranges from 300 to 349. “Proficient” ranges from 350 to 400. Students scoring above 400 are considered advanced. In Figure 1 we have indicated the Basic range through shading.

We can see that the dark line is above the lighter line on the left but crosses over and reverses toward the right. The fact the lines are not parallel represents the condition by pretest interaction specified in Table 6. A student at the low end of the ELA ability is more likely to get a boost from *History Alive!* than is a student at the higher end of ELA achievement.

Figure 2 represents the predicted difference between *History Alive!* and control groups using a single line (the dark line) to represent the distance between the two lines in Figure 1. This graph is a representation of this separation as a difference, that is, the predicted outcome for the *History Alive!* student minus the predicted outcome for a control student. Around the difference line, we provide graduated bands representing confidence intervals. These shaded bands represent how likely the difference (indicated by the dark line) could have happened just by chance (given the small sample we are working with in this experiment). These confidence intervals are an alternative way of expressing what is often called statistical significance or what we have been calling the *p* value. The dark gray band surrounding the line is the “50-50” area--the difference is as likely to be within the band as not. As we move out to the lighter bands, the difference is more and more likely to be within the lighter and lighter bands. The outer band represents conventional significance where there is only a 5% chance that a difference outside this band would have happened by chance.

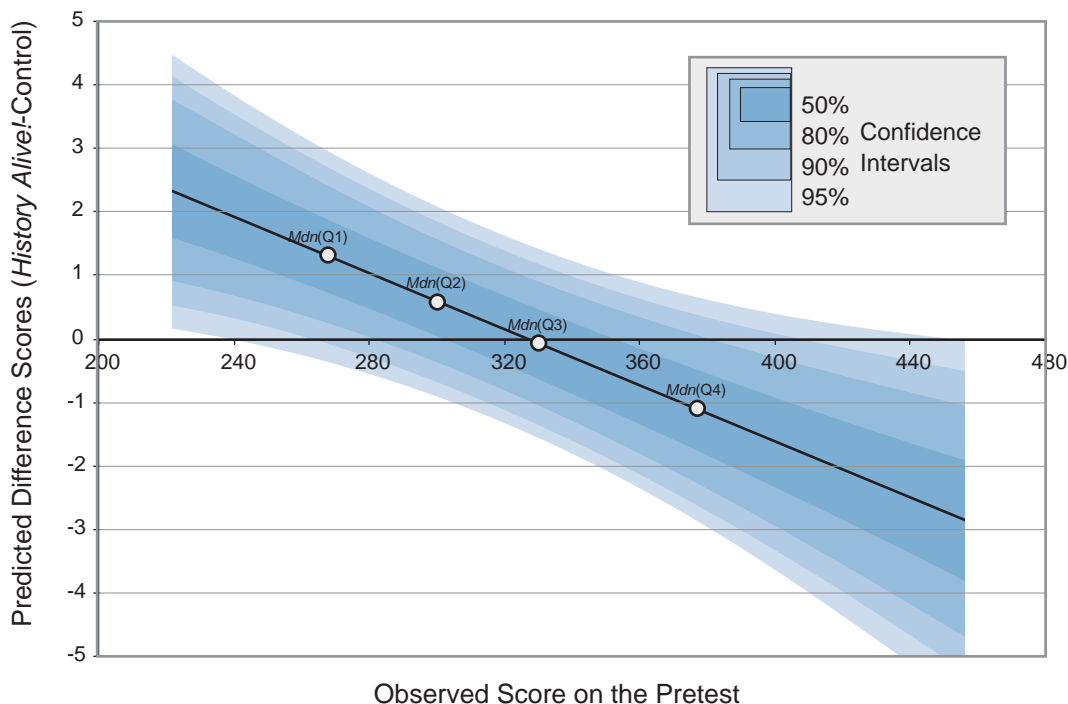
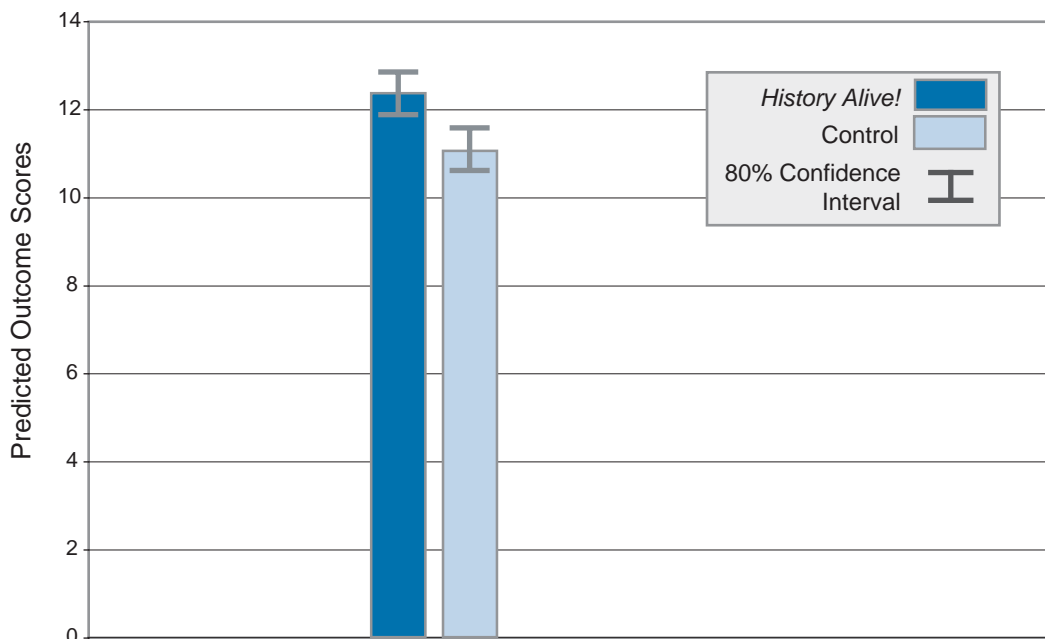


Figure 2: History outcome—difference between History Alive! and control with the predicted values for the median student at each quartile of the pretest

We see a tilt in the line and the confidence interval's rise above the axis at the lower end. We can be reasonably confident that for the students in the lower part of the English Language Arts scale, there was a measurable difference between the two conditions.

We have also indicated the location of the median student for each of the quartiles of the ELA pretest score. We are interested in the students in the lower quartile who appear to gain most from the new program.

Figure 3 shows the some of the same information as provided in Figure 2 but in bar graph form. The bar graph represents the impact of *History Alive!* for the median student in the bottom quartile of the pretest. The bar graph includes the 80% confidence interval as a marker at the top of the bars. This marker is an alternative representation of the 80% band in Figure 2. Since the markers do not overlap we have reasonable confidence that *History Alive!* would make a difference for this student.



Median Student of the Bottom Quartile of the Pretest

Figure 3: History outcome--bar graph showing the difference between History Alive! and control groups for the median student of the bottom quartile

Given the very small number of teachers involved it is possible that the assignment process to the *History Alive!* and control conditions was confounded by teachers' tendencies to focus attention on the lower achieving students. For example, by chance the four teachers in the *History Alive!* condition may have had such a tendency and the others not. This would provide an alternative explanation for the interaction. As a test of this alternative explanation, we examined the likelihood that the distribution of this tendency could have occurred by chance. Each teacher is associated with a slope such as those illustrated by the lines in Figure 1. Based on the scores just for his or her students, we can determine each teacher's slope. The slopes are then ordered and we can use a simple non-parametric statistical test to ask how likely it is that most of the control teachers would have larger values than the *History Alive!* teachers. Using the Wilcoxon rank-sum test, the p value for the ordering is 0.11 indicating an 11% chance that an ordering that clean would occur by chance. If a predisposition to focus on the lower achieving students were at play it is not highly likely that the teachers with this predisposition would mostly land in the *History Alive!* condition.

Results for History Controlling for English Learner Status

With a large population of English learners, it is relevant to the district if the *History Alive!* program would have a beneficial effect for this subgroup. In Table 7, we present a statistical model very similar to the one in Table 6 except that instead of using the prior ELA score we use the information about whether or not the student is an English learner.

Table 7: Multi-level mixed model for History—results controlling for English learner status

Descriptive statistics: CST History outcomes	Raw Group Means	Standard Deviation	Number of Students	Number of Classes	
<i>History Alive!</i>	16.074	5.326	340	11	
Control	16.442	6.448	450	17	
Mixed model: Fixed factors related to CST History outcomes	Estimate of Coefficient	Standard Error	DF	t value	p value
Intercept	14.750	0.654	26	22.56	<.0001
English learner status (non-learner = 1; learner = 0)	3.476	0.540	760	6.44	<.0001
Condition (<i>History Alive!</i> = 1; control = 0)	0.567	1.007	760	0.56	0.574
Condition by pretest interaction	-1.587	0.853	760	-1.86	0.063
Mixed model: Technical details for random components	Estimate of Variance Component	Standard Error		z value	p value
Class mean achievement	4.9016	1.6836		2.91	0.002
Within class variation	29.210	1.499		19.49	<.0001

Note: this model is based on 790 cases. Missing cases include 26 students without posttests. Another 4 cases were removed as outliers or influential points.

In this table we see just the same pattern of results. The effect of condition (whether the student was in a *History Alive!* or a control classroom) while positive was very small and the p value is relatively high indicating that this experiment could not detect a difference overall for the average student. But again we see an interaction, this time, between condition and English proficiency. When working with a dichotomous variable on the x-axis (is an English learner or not), we cannot display a scatterplot as we did in the previous analysis. We can display the results and the interaction as a bar graph.

In Figure 4, the bar graphs show the mean history scores for the English learners on the left and the fluent English speakers on the right. Each pair of bars shows the means for the two conditions. The pattern is the same as we saw in the analysis of the interaction with ELA. That is, the *History Alive!* condition favors the English learners. However, the 80% confidence interval markers in this case overlap indicating that we cannot confidently distinguish these differences from zero.

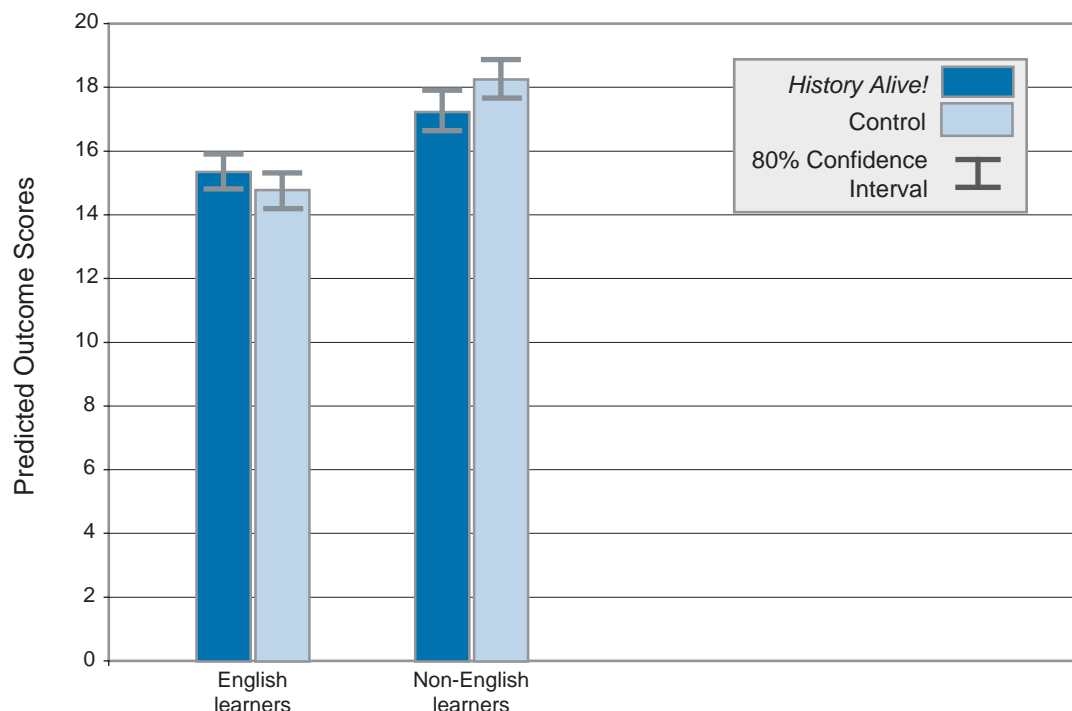


Figure 4: History outcome--bar graph showing the difference between English learners and non-English learners in both History Alive! and control groups

Results for English Language Arts

We explored the possibility that *History Alive!* may have an impact beyond the history outcome measure itself. Since we have considered ELA achievement as related to history achievement, we also modeled the ELA outcome controlling for the ELA pretest. Table 8 presents the results of the statistical model for ELA scores, using prior CST ELA scores as a covariate. This table takes the same form as tables 4 and 5 and again we see the same pattern. There is a very small positive impact of *History Alive!* for the average student but not distinguishable from zero given the large p value. The interaction is in the same direction and with a very low p value indicating that it is not very likely to be the result of chance.

Figure 5 is a scatterplot that graphs where the students fell in terms of their CST ELA pretest score (horizontal x-axis) and their CST ELA posttest score (vertical y-axis). Similar to Figure 1, the dark line is the prediction for the *History Alive!* students depending on their pretest. We see the same pattern with the lighter line for the control group crossing over the dark line, illustrating the interaction.

Table 8: Multi-level mixed model for ELA—results controlling for ELA pretest

Descriptive statistics: CST ELA outcomes	Raw Group Means	Standard Deviation	Number of Students	Number of Classes	
<i>History Alive!</i>	326.408	41.901	343	11	
Control	328.300	48.859	454	17	
Mixed model: Fixed factors related to CST ELA outcomes	Estimate of Coefficient	Standard Error	DF	t value	p value
Intercept	325.82	2.149	26	151.59	<.0001
Pretest score (centered at the mean)	0.862	0.029	607	29.65	<.0001
Condition (<i>History Alive!</i> = 1; control = 0)	2.194	3.362	607	0.65	0.514
Condition by pretest interaction	-0.141	0.050	607	-2.82	0.005
Mixed model: Technical details for random components	Estimate of Variance Component	Standard Error		z value	p value
Class mean achievement	48.413	22.162		2.18	0.015
Within class variation	604.26	34.742		17.39	<.0001

Note: this model is based on 637 cases. Missing cases include 181 students without pretests or without posttests. Another 2 cases were removed as influential points.

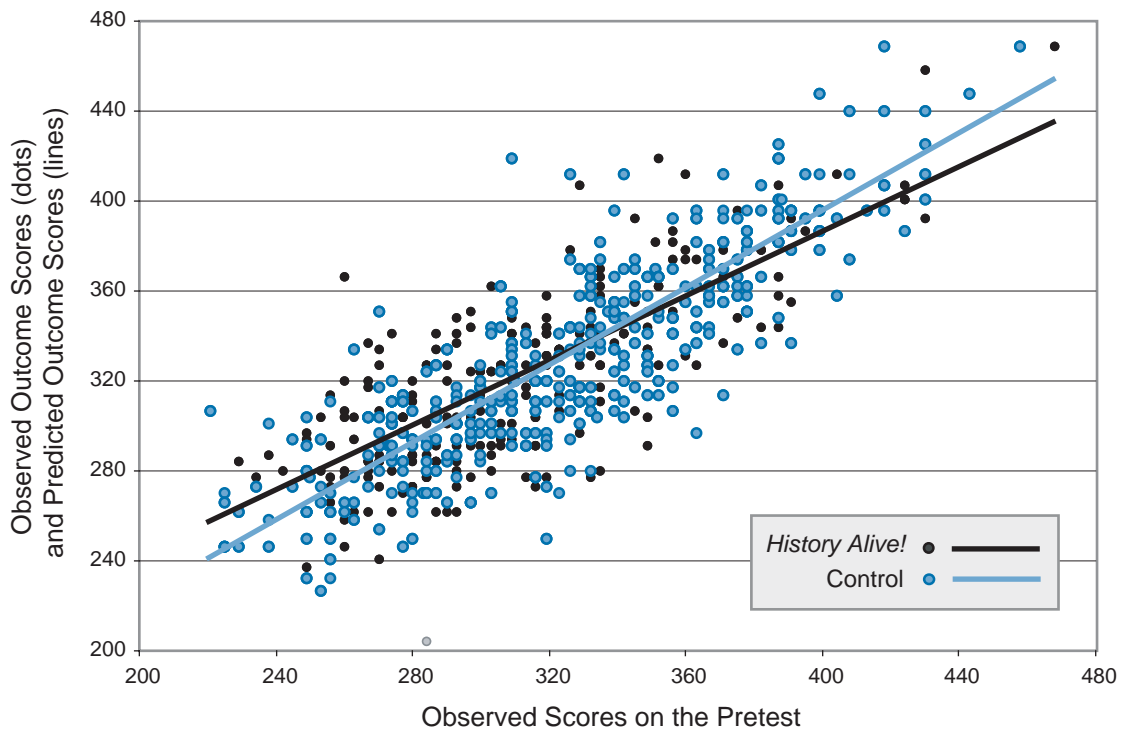


Figure 5: English Language Arts outcome--scatterplot showing History Alive! and control group students with lines showing the predicted values based on pretest score

Figure 6 represents the predicted difference between *History Alive!* and control group. As with Figure 2 the graph also indicates the locations of the median student in each of the quartiles of the pretest (ELA score from prior year).

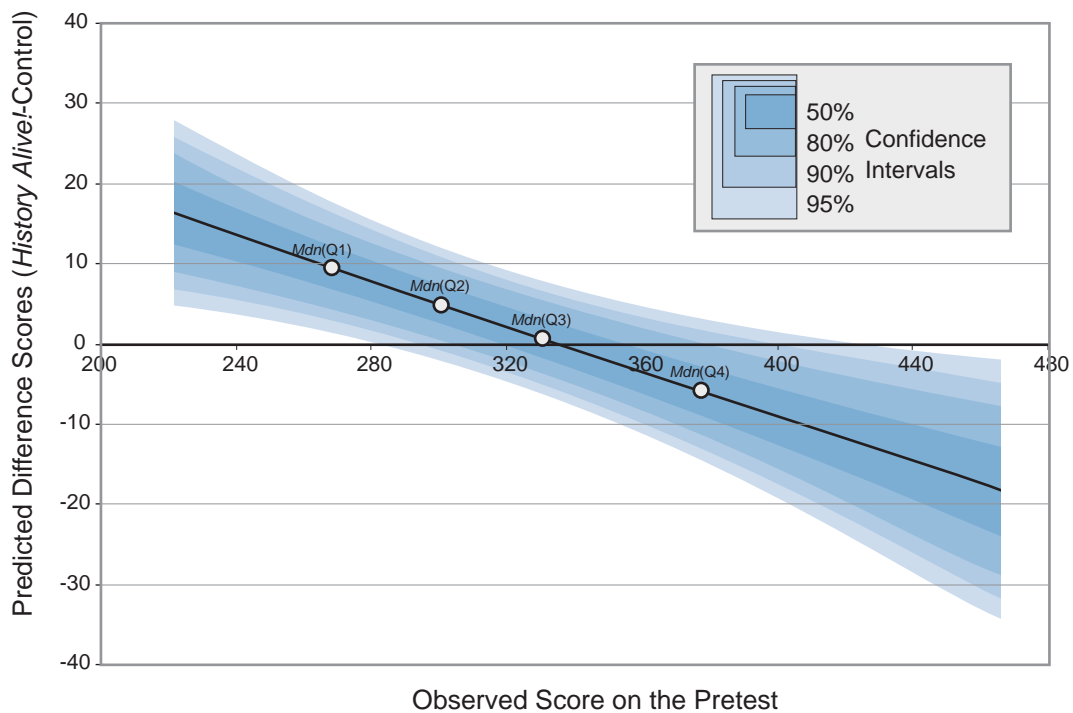


Figure 6: English Language Arts outcome—difference between *History Alive!* and control with the predicted values for the median student at each quartile of the pretest.

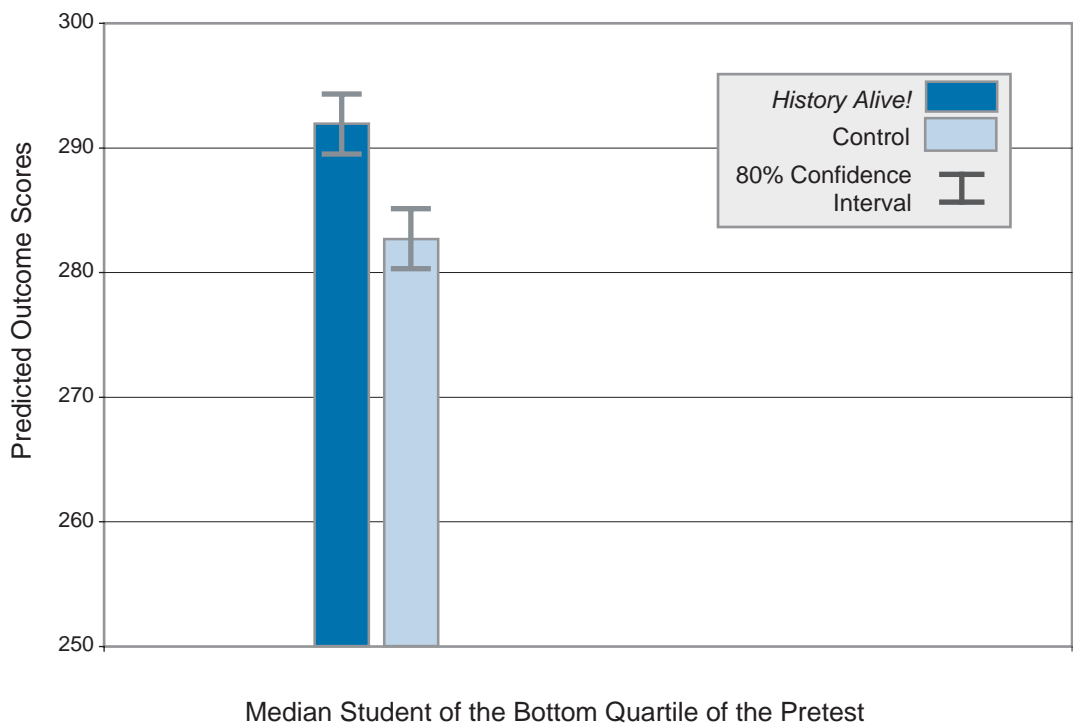


Figure 7: English Language Arts outcome—bar graph showing the difference between *History Alive!* and control group for the median student of the bottom quartile.

The pretest score for the median student is 270, a point at which point the difference line is sufficiently far from 0 that the 95% confidence interval also stays above 0. We can represent this same information as a bar graph for this median student in the bottom quartile of the pretest. Figure 7 shows the difference for that median student between being in a *History Alive!* classroom versus a control classroom.

Discussion

Our randomized experiment in this district provides evidence of a positive impact for their lower scoring students of working with TCI's *History Alive!* in comparison to what can be expected with the conventional textbook programs. Although there was no evidence of an advantage for the average student for which both programs gave similar results, we found consistently the interaction between the condition and the pretest score.

The limitations of this study must be considered in the interpretation. First, we know that to a small extent, the control group was "contaminated" in the sense that many of the control teachers had some of the TCI supplementary materials as well as training in the TCI approach. In testing in this district, it is important that the control group represent the history program as it is currently conducted. That is, if the research is to provide this district a measure of the impact of adding a new program, we have to represent the current situation, which includes a prior investment in some TCI materials and training. The prior use of the TCI materials would be expected to lower the contrast between the *History Alive!* and control groups and possibly make the apparent impact smaller than it might have been without the use of the program in the control classrooms.

Another limitation is the small number of teachers who participated. With only 9 teachers altogether, there is a danger of bias introduced by, for example, more enthusiastic teachers falling by chance in the *History Alive!* or in the control group. The coin toss guaranteed that any bias (if it exists) was not introduced intentionally. With a larger pool of teachers, the likelihood of this bias occurring would have been lower. With respect to the possibility that the condition by ELA achievement interaction was due to a similarly imbalanced distribution of teachers, we showed for ELA that it is unlikely to have occurred by chance. Ultimately the district decision-makers will have to use their local knowledge of the participants to gauge the extent of bias if any.

Our result for the English Language Arts outcome is intriguing and warrants further investigation. The result suggests that, for students at the low end of ELA achievement, an active and media rich experience in one subject may generalize to other areas. We did not collect data on the ELA classes and teachers or the extent to which the classes stayed in the same clusters for both subjects. We see the result as providing support for the conclusion that *History Alive!* is differentially effective for the lower scoring students.

Our goal in this research was to provide the participating district with evidence that would be useful in selecting among the middle school history programs that are up for adoption this year in California. While the overall difference between the programs being compared was not great, the important finding was that *History Alive!* differentially benefits the students with lower ELA scores and possibly those who are learning English. In districts with large numbers of such students, this program may have the effect of reducing the achievement gap while not reducing the achievement of the students already fluent in English or scoring well on the test of English Language Arts. Additional replications of this research in districts with different characteristics will help to expand the generality of these findings and increase our confidence in the results. Considered as a local pilot in the Alum Rock, the study adds to the information available on which to base their adoption decisions.

References

Teacher's Curriculum Institute (2006). TCI Brings Learning Alive

Retrieved on March 7, 2006 from <http://www.teachtci.com/default.asp>

Appendix A

History Study Survey 01 *History Alive! Teachers*

1. Please identify yourself.

In general, please answer the following questions for all your class periods.

2. Since the last survey until now, what percentage of class time was spent on *History Alive!* related activities?

3. Since the last survey until now, what percent of class time was spent on other history materials?

4. Since the last survey until now, indicate which chapters were addressed in your history class(es). Mark all that apply.

- | | |
|---|---|
| <input type="checkbox"/> None | <input type="checkbox"/> Mexican |
| <input type="checkbox"/> Native Americans | <input type="checkbox"/> Era of Reform |
| <input type="checkbox"/> European Explorers | <input type="checkbox"/> African Americans |
| <input type="checkbox"/> English Colonies | <input type="checkbox"/> A dividing Nation |
| <input type="checkbox"/> Life in Colonies | <input type="checkbox"/> The Civil War |
| <input type="checkbox"/> Toward Independence | <input type="checkbox"/> Reconstruction |
| <input type="checkbox"/> Declaration Independence | <input type="checkbox"/> Tensions in the West |
| <input type="checkbox"/> American Revolution | <input type="checkbox"/> Rise of Industry |
| <input type="checkbox"/> Creating Constitution | <input type="checkbox"/> Immigration |
| <input type="checkbox"/> More Perfect Union | <input type="checkbox"/> Progressive Era |
| <input type="checkbox"/> Bill of Rights | <input type="checkbox"/> World Power |
| <input type="checkbox"/> Early Republic | <input type="checkbox"/> Twenties/Depression |
| <input type="checkbox"/> Foreign Affairs | <input type="checkbox"/> World War II |
| <input type="checkbox"/> North and South | <input type="checkbox"/> The Cold War |
| <input type="checkbox"/> Andrew Jackson | <input type="checkbox"/> Civil Rights |
| <input type="checkbox"/> Manifest Destiny | <input type="checkbox"/> Contemporary |
| <input type="checkbox"/> Life in the West | |

5. Since the last survey until now, what components of the *History Alive!* Program did you use in your history class(es)?

- | | |
|---|---------------------------------------|
| <input type="checkbox"/> None | <input type="checkbox"/> Desk map |
| <input type="checkbox"/> Student book | <input type="checkbox"/> Timeline |
| <input type="checkbox"/> Student notebook | <input type="checkbox"/> Internet |
| <input type="checkbox"/> Transparencies | <input type="checkbox"/> Activities |
| <input type="checkbox"/> Placards | <input type="checkbox"/> Assessment |
| <input type="checkbox"/> CD | <input type="checkbox"/> Lesson Guide |

6. List the differences between your Class in Period 1 and your general descriptions given in questions 2-5 above.

7. List the differences between your Class in Period 2 and your general descriptions given in questions 2-5 above.

8. List the differences between your Class in Period 3 and your general descriptions given in questions 2-5 above.

9. List the differences between your Class in Period 4 and your general descriptions given in questions 2-5 above.

10. List the differences between your Class in Period 5 and your general descriptions given questions 2-5 above.

11. List the differences between your Class in Period 6 and your general descriptions given in questions 2-5 above.

12. Compared to other history programs you and your students have used in the past, how engaged were your students in each of your class periods with the *History Alive!* materials? Rate on a 5-point scale where 1 is significantly unengaged and 5 is significantly engaged. (Mark one answer only for each of your classes)

	Unengaged / Engaged		
Class in Period 1	① ② ③ ④ ⑤	N/A	
Class in Period 2	① ② ③ ④ ⑤	N/A	
Class in Period 3	① ② ③ ④ ⑤	N/A	
Class in Period 4	① ② ③ ④ ⑤	N/A	
Class in Period 5	① ② ③ ④ ⑤	N/A	
Class in Period 6	① ② ③ ④ ⑤	N/A	

13. If you used materials other than *History Alive!* since the experiment began, please list materials or activities used.

14. Did you receive training or any other kinds of support, including group meetings, in the use of *History Alive!* since the initial training?

- Yes
- No

15. If yes, please describe the amount and kind of support.

16. What else would you like to tell us about your experience working with the *History Alive!* program materials?

Appendix B

History Study Survey 01 Control Teachers

1. Please identify yourself.

In general, please answer the following questions for all your class periods.

2. What is the main textbook or other instructional program you use in class?

3. Since the last survey until now, what percent of the class time was spent on activities related to this main program?

4. What other materials have you used in class since the last survey until now?

5. Since the last survey until now, what percent of class time was spent on those other history materials?

6. If *History Alive!* were included above, what percent of class time was spent on *History Alive!* since the last survey

7. Since the last survey until now, what components of the *History Alive!* program did you use (check all that apply)

- | | |
|---|---------------------------------------|
| <input type="checkbox"/> None | <input type="checkbox"/> Desk map |
| <input type="checkbox"/> Student book | <input type="checkbox"/> Timeline |
| <input type="checkbox"/> Student notebook | <input type="checkbox"/> Internet |
| <input type="checkbox"/> Transparencies | <input type="checkbox"/> Activities |
| <input type="checkbox"/> Placards | <input type="checkbox"/> Assessment |
| <input type="checkbox"/> CD | <input type="checkbox"/> Lesson Guide |

8. List the differences between your Class in Period 1 and your general descriptions given in questions 2-5 above.

9. List the differences between your Class in Period 2 and your general descriptions given in questions 2-5 above.

10. List the differences between your Class in Period 3 and your general descriptions given in questions 2-5 above.

11. List the differences between your Class in Period 4 and your general descriptions given in questions 2-5 above.

12. List the differences between your Class in Period 5 and your general descriptions given questions 2-5 above.

13. List the differences between your Class in Period 6 and your general descriptions given in questions 2-5 above.

14. Compared to other history programs you and your students have used in the past, how engaged were your students with the materials? Rate on a 5 point scale where 1 is significantly unengaged and 5 is significantly engaged. Choose one answer only for each of your classes.

	Unengaged / Engaged	
Class in Period 1	① ② ③ ④ ⑤	N/A
Class in Period 2	① ② ③ ④ ⑤	N/A
Class in Period 3	① ② ③ ④ ⑤	N/A
Class in Period 4	① ② ③ ④ ⑤	N/A
Class in Period 5	① ② ③ ④ ⑤	N/A
Class in Period 6	① ② ③ ④ ⑤	N/A

15. What else would you like to tell us about your experience with this experiment or any other aspect of your experience that is relevant?