The project has developed a comprehensive evaluation plan, in conjunction with Drs. Matthew Russell and Rachelle Hackett of the Center for Evaluation and Research, LLC, which includes both process and summative components. The goal of the project's process evaluation component is to provide regular performance feedback to the project stakeholders to assist in the examination of the effectiveness of project implementation strategies. The process evaluation captures the forces that influence the initiative's implementation including an analysis of program elements such as context, access, services, staffing, linkages with other programs, participation levels, and collaborating agency personnel. The process evaluation involves direct observation, examination of program records, interviews with staff members and representatives to analyze the investment of the different partners, communication among agencies and organizations, and how well the different partners are working together to accomplish the goals of the program. Regular meetings between the evaluation team, project staff, participating schools, and other stakeholder groups will ensure that findings are shared and continuous improvement takes place.

The summative evaluation component seeks to understand the impact of the program services specific to teacher competencies and student achievement. The evaluation will monitor participation levels of teachers in professional development activities and the extent to which teachers integrate what they have learned into their classrooms.

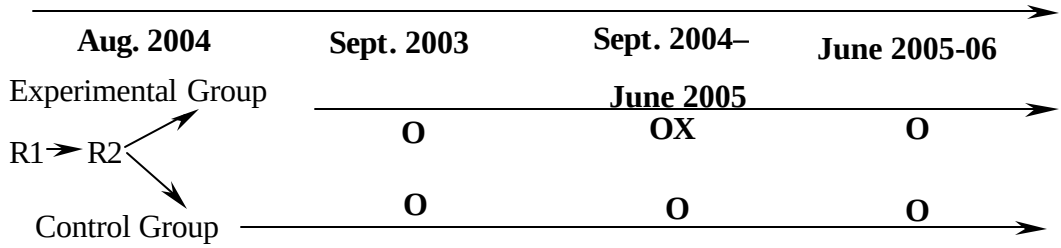
Evaluation of student achievement includes: a) an assessment of changes student performance on a standards-based assessment of social studies, b) an examination of changes in student advance placement enrollment and test performance; c) an analysis of student performance on standardized tests, and d) an examination of changes in student participation in civic duties (voting, community service etc.). The project will use within-groups and comparison group analysis of changes in standardized test scores using experimental methodology. Student achievement data will be presented both school-wide and disaggregated by the populations population categories identified in Section 1111(b)(3)(C)(xiii) of the Elementary and Secondary Education Act.

The evaluation has taken seriously the challenge of the *No Child Left Behind* statute particularly in regards to the utilization of more rigorous "scientifically based research" methods. In response to this challenge, the project has adopted a randomized experimental design for impact assessment of student outcomes beginning in August of 2004. The project will randomly

select and assign 200 high school history students to experimental (project classrooms) and control (non-project classrooms) groups. Both groups will be assessed in September of 2004 before the onset of services. These same students will be reassessed annually with the same assessments to measure initial and longitudinal impact. The following figure provides an overview of this process:

Figure ___: Project's Experimental Design Model

R1 = random selection - 200 high school students
R2 = random assignment - 100 participants in two groups; experimental and control
O = use of program measurements (see below)
X = program intervention – participation in project classrooms



O = Program Measurements
Students: California Achievement Test (CAT/6); STAR Social Science Section; History of Civil Rights Exam

The following table provides an overview of the project's evaluation questions, anticipated outcomes and the use of qualitative and quantitative analytical methods:

Table__: Evaluation Questions, Intended Outcomes and Analysis

Evaluation Questions	Intended outcomes	Analysis
1. How do teacher knowledge and qualifications change as a result of participating in the project?	Annually, 90% of the teachers and teacher aides will demonstrate increases in professional knowledge and qualifications related to working to teaching history and social science as demonstrated by teacher changes in the Quality of Personnel Measure, teacher portfolios and classroom observations of teachers.	Content analysis of teacher portfolios and teacher observation measure using NUDIST® non-linear reporting system; Pre-Post analysis (Mann-Whitney non-parametric) for QPM;
2. How do project students perform in comparison to non-project students in respect to history and social science assessments? 3. What is the longitudinal impact of the project on students' performance on history and social science assessments?	Annually, project students who participated in a project classroom in previous years will demonstrate statistically significantly ($p<.05$) higher scores in history and social science than non-participating school-age students as measured by the CAT/6 and STAR achievement tests.	Hierarchical multiple linear regression where treatment indicator (1=Treatment; 0= Control) is added after control variables (race, gender, ethnicity, SES, and language status) are in the model; Two-way mixed design ANOVA with condition (exp. vs. control) as the between Ss factor and grade (9-12) as the within-subjects factor.

In general, chi-square analyses is used for nominally-scaled data, Mann-Whitney tests for ordinal data, and multiple regression analyses where condition is dummy-coded (1= experimental; 0= control) to compare the two randomly assigned, independent groups on scaled (i.e., interval or ratio) data. For research questions involving nominal data gathered on samples that are dependent (e.g., 2c where experimental teachers only are compared between two points in time), either the McNemar or Stuart-Maxwell test are employed (i.e., for a 2x2 or 3x3

contingency table, respectively). For scaled data where means from dependent samples are compared, repeated measures ANOVAs (or their analog using the General Linear Model) are performed.

In order to reduce our error term and/or control for initial background differences between the two conditions, we will use hierarchical multiple regression where the control variables enter before the treatment indicator (1=experimental; 0=control). When statistical significance is found for the treatment indicator variable, it implies that knowing which condition the students were in (experimental vs. control) helps to predict their scores on the outcome measure being modeled. The associated change in R^2 when statistical significance is indicated depends on the degrees of freedom (which is related to the sample size and number of parameters being estimated). Therefore, we'll also consider the result to be of practical significance when it explains an additional 5% of the variability in the outcome measure beyond that accounted for by the set of control variables. Regression will be used to assess experimental effect while controlling for the identified 5 variables.

The following table provides an overview of the power analysis with R^2 at 0.05 (a very conservative calculation--designed to detect small effects with lots of variability).

Table ___ Numeric Results of Power Analysis – Independent Variables Tested/Controlled

Power	**N**	**Alpha**	**Beta**	*Count*	**R2**	**Count**	**R2**
0.90066	160	0.05000	0.09934	1	0.05000	6	0.20000

A sample size of 160 achieves 90% power to detect an R^2 of 0.05 attributed to one independent variable, the treatment indicator, using an F-Test with a significance level (alpha) of 0.05. The variables tested are adjusted for an additional six (control) variables with an R^2 of 0.20. Because this power analysis is conservative, our proposed sample size of 200 (which exceeds the

minimum just noted) should be more than adequate. Furthermore, should we find it unnecessary to control for the full set of six variables (which, for example, may include gender, race, ethnicity, SES, migrant worker status, language status, etc.), our power would be further strengthened.

In summary, the project's comprehensive evaluation approach utilizes a wealth of qualitative and quantitative information for program improvement and validation. The experimental design used to measure student outcomes meets the definition of scientifically based research, as defined in the *No Child Left Behind* statute. Specifically, the evaluation of has meet the following five criteria

1. Employs systematic, empirical methods that draw on observation and experiment;
2. Involves rigorous data analyses that are adequate to test the stated; questions and provide a justification for the general conclusions drawn;
3. Relies on measurements that are reliable and valid;
4. Utilizes an experimental design with appropriate controls; and
5. Activities are sufficiently detailed to be replicated.

The evaluation will provide compelling evidence of the project's success (*No Child Left Behind Statue,* 2002). Important policy level information regarding the teaching of American history will emerge from the findings. The project evaluators will submit a proposal outlining evaluation findings to the American Education Research Association in August of 2005.