PART B. DESCRIPTION OF STATISTICAL METHODS

B1. Respondent Universe / Sampling Methods

As indicated in Table 2, the unit of assignment in this study is the teacher. Based on our power analysis, we are planning to recruit about 120 teachers and 4,800 students (i.e., 40 students per teacher and each teacher will be randomized to treatment and control condition).

We will identify 120 teachers throughout California and Arizona that have not been exposed to PBE. We will begin by eliminating from the pool of possible teachers those who are already known to be using BIE materials. This is more likely to be an issue in California where dissemination has occurred in select districts and schools. Prior to signing up a teacher for the study, we will ask again to make sure PBE is not in use already anywhere in that school. To participate in the study, we also require study teachers to be scheduled to teach economics in the fall semester of 2007 AND the spring semester of 2008. During recruitment, we will ask for confirmation that the teacher is expecting to teach in both semesters of the academic year. This information will be confirmed with the school principal or other relevant administrator in charge of course scheduling assignments. Because of some shifting responsibilities typically experienced by teachers, we expect the number of teachers who teach economics in consecutive semesters to decline slightly. We will be interested in knowing whether, and to what extent, teachers' decisions not to teach in spring 2008 is related to the PBE intervention. In the event that it is related, it would introduce a source of non-random bias and confound a true intent-to-treat impact estimate for students in spring 2008. We believe screening for teachers who expect to teach in both semesters will solve this problem. Teachers who do not end up teaching in spring 2008 will be provided with a brief questionnaire asking for the reasons why their course schedule changed and, if in the treatment group, assessing whether the PBE curriculum had a role in the decision.

To ease recruitment burden, we will contact high schools with student enrollments of greater than 2,000 students. These schools are most likely to have teachers who teach economics in both semesters. Approximately 30 and 13 percent of high schools in California (485/1653) and Arizona (41/320), respectively, have enrollments of 2,000 or greater. Analyses of the Professional Assignment Information data collected via the California Basic Educational Data System indicates that in high schools with 2,000 or more students, an average of 3.3 teachers per school taught economics in the fall semester in California – teaching an average of 2.3 sections. We anticipate that slightly higher numbers of sections are taught in the spring, as students often take U.S. government in the fall prior to economics. As noted earlier, the study will include 120 teachers (and their students) from approximately 40 schools.

Power Estimates. For student outcomes, we assume intra-cluster correlations (ICC) of 0.15 to 0.20 based on Schochet's (2005) work and other recent work (Bloom, Richburg-Hayes, & Black, 2005; Hedges & Hedberg, 2006). Our statistical power analyses also

assume between- and within-teacher R^2 values of 0.50 for student achievement on the TEL (see Schochet, 2005), and between- and within-teacher R^2 values of 0.30 for the performance assessments, respectively. Because we have insufficient knowledge regarding the explanatory power of covariates for the teacher outcomes, we conservatively assume that covariates explain 20% of the between-teacher variation in these outcomes. This is a conservative assumption. In WestEd's *Quality Teaching for English Language Learners Study*, just a single pretest measure of pedagogical content knowledge accounted for 28 percent of the variance in this outcome. Hill and Ball's (2004) analysis of teachers' content knowledge for teaching mathematics indicated that 47 percent of the within-school variance in this outcome was accounted for by a pretest measure. Schweingruber and Nease (2000) accounted for 38 percent of the variance in teacher content knowledge in mathematics after controlling for pretest scores, teaching efficacy, and other covariates.

Table 11 shows MDES estimates for different numbers of teachers, different outcome variables, and for different ICC values. The assumptions that are the basis of the estimates are described in the table. With 120 teachers per condition and 40 students per semester instructed by each teacher, we estimate the MDES to be 0.14-0.17 for the TEL and 0.19-0.22 for the performance assessments. The MDES for teacher outcomes is 0.46. Although substantially larger than MDESs for student outcomes, teacher level MDESs of such magnitude are acceptable because impacts at the more proximal teacher level will tend to produce smaller subsequent impacts at the more distal student level. The proposed sample design thus provides adequate power to detect realistically attainable impacts on student performance and teacher outcomes.

Table 11. Minimum Detectable Effect Size by Number of Teachers and ICC

Total Schools	Total Teachers	*Test of Economic Literacy (40 Students per Teacher)* ICC		*Performance Assessment (16 Students per Teacher)* ICC		*Teacher Outcomes* ICC
		0.15	**0.20**	**0.15**	**0.20**	
30	90	0.18	0.20	0.23	0.25	0.53
40	120	0.15	0.17	0.19	0.22	0.46
50	150	0.14	0.15	0.17	0.19	0.41

Notes:

Calculations are based on the following assumptions: (a) 3 high school economics teachers per school, (b) high school economics instructors teach 2 sections in a particular semester; (c) 25 students are enrolled in each economics class, and 80% of those students (20) will provide valid outcome data; (d) balanced allocation between treatment and control condition, (e) statistical power levels of .80, (f) Type I error rates of .05 (two-sided), (g) a fixed-effects statistical model, (h) the covariates used in the analysis explain 50% of between- and within-teacher variance in student scores on the Test of Economics Literacy, (i) the covariates used in the analysis explain 30% of between- and within-teacher variance in student scores on the Performance Assessments, and (j) the covariates used in the analysis explain 20% of the between-teacher variance in teacher outcomes. Shaded cells correspond to expected sample size.

B2. Data Collection Procedure

Detailed data collection procedure and timeline is displayed in Table 4.

B3. Methods to Maximize Response Rates and to Deal with Non-response Issues

Assuring High Response Rates. We have begun recruitment by contacting superintendents who are ready to consider conducting a large-scale research study on economics in their district. Data structures have been developed to track recruited districts, carefully acknowledging that once a district or school has expressed interest to be in the study, it is possible that they will drop out of the study at a later date. The recruitment team makes an initial call to the district superintendent's office and asks to be connected with the Director of Curriculum and Instruction for the district. Further inquiry then connects district and school-level staff, resulting in an initial assessment of interest. If the exploratory conversations continue to generate interest from site principals and then economics teachers, the recruitment team will update the Director of Curriculum and Instruction. It is our expectation that the participation by a teacher in this study will ultimately require complete review by the superintendent and school board before a formal agreement can be executed.

To the extent possible, the final sample of teachers, within the schools that they teach, will be representative of California and Arizona in terms of gender, race/ethnicity, socioeconomic status, and geographic location. For teachers to be eligible, a prescreening interview will be conducted with the research team to establish the availability of individualized data from the standards-based student-level assessment in each of the states. To the extent needed, schools will refer the research team to district information systems specialists to review data extraction at the student level for these data sources.

Once oral confirmation of study participation is received, a memorandum of understanding (MOU) will be sent to each teacher and school outlining the support they will receive for participating in the study, the roles and responsibilities of both research staff and school site staff, and estimates of the time required to collect data. The formal letters and permissions will also include language that signals an agreement not to use the curriculum if assigned to the control group: this alleviates the concern about adoption of the curriculum by the control group through online download (without professional development) after learning about the curriculum.

We will use a combination of good survey design, good initial collection of contact information, and very persistent follow-up to achieve high response rates. Survey data are processed immediately to identify non-respondents, who are then scheduled for follow-up administration. Teachers will be compensated for their time spent during the summer 2007 professional development seminar, academic year online seminars, and completing survey instruments. In our experience, it is most important to closely monitor the progress of survey administration and to make quick and decisive adjustments to the survey protocol when response rates fall below key targets. Such flexibility requires high-

level attention to survey progress. Extensive experience in administering and managing survey efforts enables us to recognize when problems occur and to take steps to address them.

No Shows. Although this study includes a plan to monitor and ensure implementation fidelity, it is possible that non-trivial numbers of participants assigned to the treatment group will not participate in all intervention activities. Nonparticipation by significant numbers of those targeted to receive the intervention would likely dilute potential program impacts. Extensive efforts will be made to collect data from such non-participants, and levels of participation in the intervention will be monitored through surveys and records. So as not to bias impact estimates, all such participants will be kept in the impact analysis in their original, assigned groups to avoid sample selection bias. That is, an intention-to-treat analysis (ITT) will be performed. ITT refers to the fact that random assignment only establishes an "intention to treat," but does not actually guarantee that those assigned to the program experience it.

Attrition. With the proposed design, student attrition is likely to be infrequent over a semester, and teacher attrition is unlikely in the middle of an academic year. Even though the design mitigates the chances of attrition, a high level of sample attrition would be unacceptable for the integrity of the experimental design. Sample attrition relates to our ability to collect outcome data on all who were randomly assigned at the start of the study. Serious violations in this regard will likely cause significant biases in the estimated program effects. There is no reliable way to identify control teachers to accompany the program teachers that left the study. For this reason, it is critical that any teacher who agrees to participate in this study remain involved in the research efforts until all data collection is completed, even if they were unable to fully implement the intended program treatment. This is a key focus of our upfront recruitment efforts.

Although all efforts will be made to minimize attrition from the study, our estimates of treatment effectiveness will be biased to the extent that unmeasured factors associated with attrition are related to predictor and outcome measures. To correct for this potential bias, we will use Heckman's (1979) two-stage estimator to "partial out" the association between non-random attrition and our outcome variables. This method is similar to the propensity score method developed in the prevention literature (Rosenbaum & Rubin, 1983, 1984). We will also experiment with multiple-imputation techniques to impute values for respondents who dropped out of the study (Schafer, 1997).

Both of these strategies – sample selection model estimation and multiple-imputation techniques will only be used as a last resort if there is evidence of data values being missing in a non-random manner. By far, the best method of handling this type of bias is to reduce the probability that attrition will occur in the first place. With the current proposed design – with one year of implementation targeting economics teachers who are scheduled to teach for two subsequent semesters – we feel that sample attrition is unlikely to occur with substantial frequency, and is even more unlikely to be driven by treatment status. Examining baseline differences between teachers that drop out of the study, by treatment and control condition, is likely to yield the most valuable information with regards to this issue.

B4. Expert Review and Piloting of Instruments

All data collection instruments will go through a series of reviews before being used in the field. The research team will work closely to ensure that the instruments are reliable and valid. Instruments and procedures will also be shared with the TWG members, who will bring their expertise as researchers and practitioners to bear on the design of the items, the burden on respondents, and the implications for data analysis.

In addition to such review, the research team will pilot test the following instruments with a small group of people since they are either newly developed or are developed based on some existing instruments but have not yet administered before:

- Teacher Background Survey / Teacher End of Semester Survey[1]
- Student Background Survey / Student End of Semester Survey[2]
- Teacher End of Unit Surveys

During these piloting, which will be administered to no more than six respondents, the team will assess item comprehension, the effectiveness of the proposed strategies for gaining cooperation, and the length of time for respondents to answer questions in the instruments. Such information is critical for determining the burden associated with each instrument, which must be presented to all respondents before the administration of any federally sponsored research questionnaire to more than nine respondents.

B5. Statistical Consultants

Neal Finkelstein, PhD, serves as the Project Director for this study. Dr. Finkelstein is currently the Co-director for Research Studies for the Regional Educational Laboratory (West) (REL West). As a Senior Research Scientist, he develops research and evaluation designs that study the impact of program implementation in K–12 public schools. He ensures that evaluation designs feature high standards of evidence, and oversees the implementation of randomized field trials in education settings, including site recruitment, data collection and analysis.

Prior to WestEd, Finkelstein worked on large-scale program evaluations and policy analyses encompassing K–12 and higher education, and the bridge between them. His areas of expertise include K–12 school finance, academic preparation programs for high school youth, school-to-work, and early childhood education. Each area involves the collection, management, and analysis of large quantitative data sets as well as questions of cost, cost-effectiveness, and the marginal cost of policy decisions in education at the state and federal level.

---

[1] Both surveys contain some common core measures, so only one of them will need to be field tested.
[2] Both surveys contain some common core measures, so only one of them will need to be field tested.

Finkelstein served as Director of Educational Outreach Research and Evaluation for the University of California Office of the President. There he implemented research and evaluation designs that studied the effectiveness of K–12 student and school academic programs initiated by the University of California on 10 campuses throughout the state. These programs emphasized the connections between K–12 education and postsecondary education opportunities for students.

Finkelstein can be reached by phone at (877) 938-3400, ext. 3171.

Chun-Wei (Kevin) Huang, PhD, serves as a Senior Data Analyst responsible for instrument design and data analysis for this study. As a Senior Research Analyst at WestEd, he works with other researchers to design and implement rigorous experimental trials within WestEd's Regional Educational Laboratory (West) (REL West). He ensures that the instruments used in these studies are reliable and valid and is responsible for conducting statistical analyses during all phases of the research. In addition to his work with REL West, he provides assistance to colleagues with statistical and measurement modeling for other WestEd projects.

Prior to WestEd, Huang worked at CTB/McGraw-Hill as a Research Scientist. He was involved in several projects including two statewide testing programs. His main responsibilities as a Research Project Manager were to lead and conduct data analyses (e.g., test equating and scaling) in accordance with customers' requirements. He has taught statistics at both the undergraduate and graduate level.

Huang can be reached by phone at (877) 938-3400, ext. 3162.

REFERENCES

Baker, E. L. (1997). Model-based performance assessment. *Theory Into Practice, 36*, 247-254.

Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and Cognition* (pp. 131-153). Englewood Cliffs, NJ: Prentice-Hall.

Baker, E. L., Linn, R. L., Abedi, J., & Niemi, D. (1996). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research, 89*, 197-205.

Baker, E. L., & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*, 269-282.

Bloom, H.S., Richburg-Hayes, L., & Black, A.R. (2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions*. New York: MDRC.

Bryk, A. S., & Driscoll, M. E. (1988). *The high school as community: Contextual influences and consequences for students and teachers*. Madison, WI: National Center on Effective Secondary Schools, University of Wisconsin.

Expeditionary Learning Outward Bound. (1999). *Early indicators from schools implementing New American Schools designs.* Cambridge, MA: Author.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Oxford University Press.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*, 153-161.

Hedges, L.V. (2006). Intraclass correlation values for planning group randomized trials in education. Institute for Policy Research Working Paper WP-06-12. Evanston, IL: Northwestern University.

Hendrie, C. (2003, April 23). Small schools hard to start, report finds. *Education Week, 22*(32). Retrieved from http://www.edweek.org/ew/articles/2003/04/23/32gates.h22.html

Hill, H.C. and Ball, D.L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education.* 35(5):330-351.

Hodgin, R. F. (1984). *Attitude assessment for research in economics education*. (ERIC Document Reproduction Service No. ED 248 779).

Honey, M., & Henríquez, A. (1996, April). *Union City interactive multimedia education trial: 1993-95 summary report* (CCT Reports, Issue No. 3). Retrieved from http://www.edc.org/CCT/ccthome/tech_rept/CCTR3/

Kim, S., & Wilson, M. (2006). *Analysis of a high school economics test: Food court*. Unpublished report. University of California, Berkeley.

Mergendoller, J., Maxwell, N., & Bellisimo, Y. (in press). The effectiveness of problem-based instruction: A comparative study of instructional methods and student characteristics. To appear in *Interdisciplinary Journal of Problem Based Learning.* Retrieved from http://www.bie.org/files/IJPBL%20PBE%20PaperFINAL-single%20spaced.pdf

Mo, K., & Choi, Y. (2003). Comparing problem-based learning with traditional instruction: Focus on high school economics. *Theory and Research in Citizenship Education, 35*(1), 89-113. Published by the Association of Social Education in Korea (ISSN: 1598-7280). English abstract is available: http://www.bie.org/research/pbss/econ/summary.php?id=39

Moeller, B. (2005, April). Understanding the implementation of problem-based learning in New York City high school economics classrooms. In J. Ravitz (Chair), *Assessing Implementation and Impacts of PBL in Diverse K-12 Classrooms*. Papers presented at the annual meeting of the American Educational Research Association, Montreal, Canada. Retrieved from http://www.bie.org/AERA2005/Moeller_Paper.pdf

Murray, D. M. (1998). *Design and analysis of group randomized trials*. New York: Oxford University Press.

National Council on Economic Education (NCEE). (1999). *Standards in economics: Survey of students and the public*. New York: Author. Retrieved from http://ncee.net/cel/results.php

National Council on Economic Education (NCEE). (2003). *Survey of the states: Economic and personal finance education in our nation's schools in 2002.* New York: Author. Retrieved from http://www.ncee.net/about/survey2002/

Newmann, F., & Wehlage, G. (1995). *Successful school restructuring: A report to the public and educators by the Center on Organization and Restructuring of Schools*. Madison, WI: Wisconsin Center for Education Research. (Distributed jointly by the American Federation of Teachers, Association for Supervision and Curriculum Development, National Association of Elementary School Principals, and The National Association of Secondary School Principals). Retrieved from http://llanes.panam.edu/journal/library/Vol1No1/success.html

Niemi, D. (1996). Assessing conceptual understanding in mathematics: Representation, problem solutions, justifications, and explanations. *Journal of Educational Research, 89,* 351-363.

O'Neil, H. F., Jr. (1999). Perspectives on computer-based performance assessment of problem solving: Editor's introduction. *Computers in Human Behavior, 15,* 255-268.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Ravitz, J., Becker, H., & Wong, Y. (2000). Report #4: Constructivist compatible beliefs and practices among U.S. teachers. Teaching, Learning & Computing, 1998: UC, Irvine. Retrieved from http://www.crito.uci.edu/TLC/FINDINGS/REPORT4/REPORT4.PDF

Ravitz, J., & Mergendoller, J. (2005, April). Evaluating implementation and impacts of problem based economics in U.S. high schools**.** In Ravitz, J. (Chair). *Assessing Implementation and Impacts of PBL in Diverse K-12 Classrooms*. Montreal, Canada. Retrieved from http://www.bie.org/AERA2005/Ravitz_Mergendoller.pdf

Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenbaum, P., & Rubin, D. B. (1984). Reducing bias in observational studies using sub-classification on the propensity score. *Journal of the American Statistical Association, 79,* 516-524.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research.

Schweingruber, H.A. & Nease, A. A. (2000). Teachers reasons for participating in professional development programs: Do they impact program outcomes? Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April, 2000.

Trochim, W. M. K. (2001). *The Research Methods Knowledge Base*, Cincinnati: Atomic Dog Publishing (http://atomicdogpublishing.com)

Walstad, W., & Rebeck, K. (2001). *Test of economic literacy, examiners manual* (3rd ed.). New York: National Council on Economic Education.

Willison, S., & Kelly, P. (2004, April). *Improving K-12 student learning through long-term teacher professional development*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.