

PART B. DESCRIPTION OF STATISTICAL METHODS

B1. Respondent Universe / Sampling Methods

As indicated in **Exhibit 3**, the unit of assignment in this study is the school. Based on our power analysis, we are planning to recruit about 50 schools and 15,000 students (i.e., 3,000 students per school and each school will be randomized to treatment and control condition).

Sampling and Power Estimates. In order to determine the appropriate sample sizes required for the study, we calculated *minimum detectible effect sizes* (MDES – see Bloom 1995) based on the unit of randomization, the level of clustering, the availability of baseline explanatory variables, and other design characteristics using the procedures described by Donner and Klar (2000), Murray (1998), Raudenbush (1997), and Schochet (2005).

As mentioned above, 50 schools will be randomly assigned to two conditions, with three teachers/classes per grade in each school and 25 students per class. We conservatively assume a student attrition rate of about 25% for power estimation purposes, leaving 18 students per class at the end of the 2nd implementation year available for analysis of the outcomes assessed with surveys and school records. SSRS data will be available for a minimum of 9 students per class. For the purposes of the power analyses, we assume intraclass correlations of 0.15 and 0.07 for the academic and nonacademic outcomes, respectively, based on Schochet's (2005) recent work. Our statistical power analyses also assume between- and within-school R^2 values of 0.50 (Schochet, 2005). With 25 schools per condition and a minimum of 56 ($25 \times 3 \times .75$) students per grade in each school, we estimate the MDES to be 0.23 for academic outcomes and 0.17 for behavioral and attitudinal outcomes. With as few as 10 students per school, the MDES only rises to 0.28 and 0.23 for academic and non-academic outcomes, respectively — suggesting that more than adequate power is available for conducting analyses of student subgroups. For outcomes assessed via teacher and parent reports on the SSRS (27 students per grade per school), we estimate an MDES of 0.19. Precision is enhanced even more if we pool students across grades. These estimates are purposively conservative, as we do not take into account sample stratification prior to random assignment (see below) in our statistical power estimates.

A Priori Stratification. To improve precision of impact estimates and to guard against chance non-equivalence between randomly assigned conditions, schools will be blocked prior to randomization. With potentially high levels of heterogeneity across sites, non-equivalence between conditions is still possible, especially when the number of groups is limited. Similar schools will be stratified into groups based on two factors:

- 1) a composite index representing a school's socio-demographic composition; and
- 2) a school's true academic performance, holding constant the socio-demographic characteristics of its students.

The composite index of the socio-demographic composition of each school will be calculated based on school enrollment, race/ethnic composition, and the percentage of

students eligible for subsidized meals. In calculating this composite index, these factors will be weighted proportionately to how strongly they are related to a school's average academic performance (see California Department of Education, 2000). In addition, to be considered a match, a school will be required to be in the same geographic area – preferably in the same district or a neighboring district. Each potential participating school will be located in multidimensional space defined by these factors, and matched with 8-10 other schools. Within each block, schools will be randomly assigned to treatment and control conditions. We anticipate forming 5 to 7 groups of schools. Based on the TWG's recommendations, this procedure was chosen instead of pair-wise matching to preserve degrees of freedom in our analytic models. A covariate for block membership will be included in the impact analysis models.

Recruitment and Assignment to Condition. Site recruitment will take place in California using established WestEd marketing channels to identify interested schools. Staff in WestEd's Health and Human Development Program (HHDP) have extensive contacts with schools and districts interested in character education through its management of the Title IV and California Healthy Kids Survey listserv and presentations at state Title IV county coordinator meetings. Several school districts have already contacted HHDP about participation in the study.

A listserv message will be sent to all school districts in California serving students in grades 1 through 5 and to all Title IV Coordinators in early October notifying them about the study and requesting participation from schools within their respective districts and counties. Responses to the listserv message will be reviewed to ensure schools meet the criteria for selection into the study as discussed above. Follow-up messages will be sent until there is a pool of schools from which we can recruit into the study. District/county offices will be asked to provide contact information for each eligible school in their jurisdiction, and WestEd will contact each school directly to recruit them into the study. An analogous procedure will be used in Arizona.

For schools to be eligible, a prescreening interview will be conducted with the evaluation team to establish the availability of individual-level data on school grades, attendance, and state achievement tests; to screen-out schools already implementing LIC; to ensure that school staff are fully aware of the requirements of participating in randomized trial; and to ensure that sufficient numbers of teachers are willing to participate in the study. To the extent needed, schools will refer the evaluation team to district information systems specialists to review data extraction at the student level for these data sources. During the pre-screening interview, school staff will be provided an overview of the LIC program, the evaluation, and specific information about the process of obtaining parental consent. Follow-up interviews will be conducted as needed during the recruitment process.

The school sampling frame will be restricted to schools serving students in grades 1 through 5 with total enrollments of less than 800 students. Elementary schools with larger enrollments – which represent 4% of elementary schools in California – are excluded because of the expense of obtaining curriculum materials for such a large number of students and because school-level implementation is likely to be different in such large schools. We also require that at least two-thirds of teachers within each grade agree to participate in the study for the school to be included in the sample.

Once oral confirmation of study participation is received, a memorandum of understanding (MOU) will be sent to each site outlining what support and possible compensation sites will receive for participating in the study, the roles and responsibilities of both research staff and site staff, and estimates of the time required to collect data. Evaluation staff will maintain monthly contact with schools to ensure project implementation, review data collection schedule, and maintain good working relationships with school staff to minimize attrition rates. At the time of recruitment, school staff will be interviewed to determine the best timeframe for the professional development trainings. Regional trainings will be scheduled to meet the needs of school staff to ensure high levels of participation in a one-day training. The MOU will include an agreement that all teachers involved will attend the training.

In spring Year 0 (2007 for cohort 1, 2008 for cohort 2), each participating school will submit a roster of all students in grades 1 thru 4. Within each grade, a 50% random sample of students (approximately 12 per class) with positive parental consent will be selected to be focal students for teacher- and parent reports of student well-being. The process will be repeated for 1st graders in spring of Year 1.

After all recruitment activities have been completed, parental consent forms received, and baseline data collected, schools will be randomly assigned to use LIC, at no cost, in Grades 2-5 (Group #1) or to a treatment-as-usual control group (Group #2) using IES guidelines and a computerized algorithm. As noted above, schools will be stratified into groups of 8-10 prior to assignment to condition to guard against chance baseline non-equivalences and to improve precision.

B2. Data Collection Procedure

Detailed data collection procedure and timeline is discussed in A2 and summarized in **Exhibit 5**.

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

Assuring High Response Rates. We expect a student non-response/attrition rate of less than 25 percent. 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 across all the studies (also refer to A2, “Data Collection Procedure and Timeline” for specific follow-up procedures used in this study). Survey data are processed immediately to identify non-respondents, who are then scheduled for follow-up administration. We will implement special procedures to follow up with students who either move to a different school or move out of the district. We use financial or other incentives to ensure high response rates among respondents. 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 personal experience in administering and managing survey efforts enables us to recognize when problems occur and to take steps to address them. All of our proposed senior-level project staff have hands-on experience with managing survey efforts and the use of survey data in experimental and quasi-experimental studies.

No Shows. Although the study includes a plan to monitor implementation fidelity, it is possible that non-trivial numbers of teachers in schools assigned to the treatment group will not participate in intervention activities. Non-participation by significant numbers of those targeted to receive the intervention would likely dilute potential program impacts. Data will be collected 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.

Attrition. A high level of sample attrition is 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. For example, if several program group schools drop out of the study, it is likely that both the background characteristics and the expected outcomes in these schools are different from the ones that remain. As a result, program impacts may appear more or less favorable than they should. There is no reliable way to identify control schools to accompany the program schools that left the study. For this reason, it is critical that any schools that agree to participate in the proposed studies 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 the upfront recruitment efforts that are part of the study proposed here.

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).

B4. Expert Review of Instruments

All data collection instruments had been used in other studies and have been shown to possess good psychometric properties (see **Exhibit 4**). No further expert reviews or piloting would be needed for this study.

B5. Statistical Consultants

Thomas L. Hanson, PhD, is a Senior Research Associate in the Health and Human Development Program at WestEd and Co-Director of research of WestEd's Regional Educational Laboratory (West) (REL West). He directs the *Lessons in Character Outcome Evaluation* (REL West/Ed-IES) and the *Tribes Outcome Evaluation* (NIJ). Hanson also serves as lead methodologist for the *Algebraic Interventions for Measured Achievement* (ED/IES), an experimental trial testing the efficacy of an intervention curriculum targeting specific algebraic learning trouble spots; *Math Pathways and Pitfalls Lessons for K-7 Students* (ED/IES), a cluster-randomized trial investigating the efficacy of the Math Pathways and Pitfalls instructional materials on 4th-6th grade students' mathematics achievement and mathematical language development; and the *Integrating Literacy and Science Instruction in High School Biology Project* (NSF) and *Efficacy of Reading Apprenticeship Professional Development for High School History and Science Teaching and Learning* (ED-IES) studies, which are cluster-randomized trials that examine the effectiveness of teacher training in the integration of reading instruction and subject area content on student achievement in science, history, and reading.

Hanson can be reached by phone at (562) 799-5170.

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

- Anderman, L. H. (1999). Classroom goal orientation, school belonging, and social goals as predictors of students' positive and negative affect following the transition to middle school. *Journal of Research and Development in Education*, 32, 89-103.
- Bloom, H.S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, 19(5), 547-556.
- Bowen, N. K., & Bowen, G. L. (1999). Effects of crime and violence in neighborhoods and schools on the school behavior and performance of adolescents. *Journal of Adolescent Research*, 14(3), 219-341.
- California Department of Education. (2000). *Construction of California's 1999 school characteristics index and similar schools ranks*. Sacramento, CA: Office of Policy and Evaluation, California Department of Education.
- Characterplus. (2002). *Evaluation resource guide: Tools and strategies for evaluating a character education program*. St. Louis, MO: Author.
- Connell, J. P., & Halpern-Felsher, B. L. (1997). How neighborhoods affect education outcomes in middle childhood and adolescence: Conceptual issues and an empirical example. In J. Brooks-Gunn, G. J. Duncan, & J. L. Aber, (Eds.), *Neighborhood poverty: Context and consequences for children* (pp. 174-199). New York: Russell Sage.
- Dietsch, B, Bayha J., & Zheng, H. (2005). *Short-term effects of a literature-based character education program among fourth grade students*. Paper presented at the 2005 Meetings of the American Educational Research Association, Montreal, Quebec, Canada.
- Donner, A. N. & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.
- Elliott, S. N. (1995). *The responsive classroom approach: Its effectiveness and acceptability* (Final Evaluation Report). Madison, WI: University of Wisconsin.
- Funk, J. et al., (2003) The Attitudes Toward Violence Scale: Child Version. *Journal of Interpersonal Violence*, 18: 186-196.
- Furrer, C., and Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95, 148-162
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Oxford University Press.

- Gresham, F.M. and Elliott, S.N. (1990). *Social Skills Rating System (SSRS)*. Bloomington, MN: Pearson Assessments.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153–161.
- Hanson, T. L., Austin, G. A., & Lee-Bayha, J. (2004). *Ensuring that no child is left behind: How are student health risks & resilience related to the academic progress of schools*. San Francisco: WestEd. Available: www.wested.org/hks
- Kisker, E, Kalb, L, Miller, M., Sprachman, S., Carey, N, Schochet, P, & James-Burdumy, S. (2004). Social and character development research program evaluation: Supporting statement for request for OMB approval of SACD evaluation. Princeton, NJ: Mathematica Policy Research.
- Lochman, J. E., Lampron, L. B., Gemmer, T. C., & Harris, S. R. (1987). Anger coping intervention with aggressive children: A guide to implementation in school settings. In P. A. Keller & S. R. Heyman (Eds.), *Innovations in clinical practice: A source book*, 6, 339-356.
- Loeber, R. & Dishion, T.J. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin*, 94, 325-382.
- McMorris, B.J., Clements, J., Evans-Whipp, T., Gangnes, D., Bond, L., Roubourou, J.W. & Catalano, R. (2004). A comparison of methods to obtain active parental consent for an international student survey. *Evaluation Review*, 28(1) 64-83.
- Murdock, T. B., Anderman, L. H., & Hodge, S. A. (2000). Middle-grade predictors of students' motivation and behavior in high school. *Journal of Adolescent Research*, 15, 327-352.
- Murray, D. M. (1998). *Design and analysis of group randomized trials*. New York: Oxford University Press.
- Opinas, P. and Frankowski, R. (2001). The aggression scale: A self-report measure of aggressive behavior for young adolescents. *Journal of Early Adolescence*, 21(1), 50-67.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design in cluster randomized trials. *Psychological Methods*, 2(2), 173–185.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J. Beuhring, T., Sieving, R. E., Shew, M., Ireland, M., Bearinger, L. H., & Udry, J. R. (1997). Protecting adolescents from harm: Findings from the national longitudinal study on adolescent health. *Journal of the American Medical Association*, 278 (10), 823-832.

- 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 subclassification on the propensity score. *Journal of the American Statistical Association*, 79, 516-524.
- Ryan, A.M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38, 437-460.
- 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.
- Trochim, W. M. K. (2001). *The Research Methods Knowledge Base*, Cincinnati: Atomic Dog Publishing (<http://atomicdogpublishing.com>).
- Wentzel, K.R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, 89, 411-419.
- Williams, M. (2000). Models of character education: Perspectives and developmental issues. *Journal of Humanistic Counseling, Education & Development*, 39(1), 32-41.