SUBMITTED TO

Matt Dozier President/Chief Executive Officer EAST, Inc.

SUBMITTED BY

Jonathan Tunik, M.P.A., Senior Associate Lori Ramsey, Ed.D., Senior Research Associate Alan J. Simon, Ph.D., Senior Vice President

Arkansas Environmental and Spatial Technology Initiative

James Boardman

Assistant Commissioner

Division of Research & Technology

Arkansas Department of Education

State-wide Evaluation 2003-2006 Final Report



90 Broad Street Suite 1200 New York, New York 10004 212-425-8833 www.metisassoc.com



Arkansas Environmental and Spatial Technology Initiative

State-wide Evaluation 2003-2006 Final Report

December 31, 2007

Principal Investigator: Alan J. Simon, Ph.D., Senior Vice President

Project Director/Primary Author: Jonathan Tunik, M.P.A., Senior Associate

Technology Advisor, Qualitative Research Design and Report Writing: Lori Ramsey, Ed.D., Senior Research Associate

Data Analysis and Design Consultation: Julia Alemany, M.S., Research Associate Eden Nagler, M.Phil., Research Analyst Michael Scuello, B.A., Senior Associate for Design and Analysis Richard Pargament, Ph.D., President Emeritus

> Data Management: Lawrence Chau, B.B.A., Systems Associate Hugo Coello, B.S.T., Systems Associate

We would like to dedicate this report to Warren Stewart, whose wise counsel, dedication and boundless good humor contributed so much to this project; and to P.G. Bradford, whose generous nature and good cheer never failed to light up the room. They are greatly missed.

Metis Associates 90 Broad Street, Suite 1200 NY, NY 10004 212-425-8833 212-480-2176 ó fax www.metisassoc.com

Project Description and Population Served

The Environmental and Spatial Technology (EAST) Initiative is a performance-based learning environment utilizing community service, project-based, service learning, integrated with advanced technological applications in an interdisciplinary environment where the intellectual and problem-solving growth of students is the focus. The project currently serves 190 schools in eight states (Arkansas, California, Hawaii, Illinois, Iowa, Louisiana, Mississippi and Pennsylvania). At the outset of the current study (school year 2003-2004), Arkansas had approximately 130 EAST programs. The majority of EAST sites in Arkansas were in high schools (90%), with 7% at middle schools, and another three EAST programs that were implemented in an elementary school or college. Sites are spread over the predominantly rural state of Arkansas, with the majority in rural counties and the remainder divided roughly evenly between suburban and urban counties.

EAST¢s central concept is based on the importance of studentsø responsibility for their own learning, with a focus on cooperative learning, interdependence, and individual accountability, and development of problem solving, decision-making and higher-order thinking skills. The program is an elective that is intended to serve a diverse group of students, with diversity understood to include aptitude, academic motivation, and achievement levels, as well as race, gender, and economic background. Students with special needs, gifted students, at-risk students, and all students in between are placed on a level playing field by being in the same class. Ultimately, the goal should be to construct classes that are representative of the demographics of the school as a whole.

The instructional model relies on teachers (known as õfacilitatorsö) to guide and assist students as they pursue their projects, through which students connect with peers, faculty, community members, and business and university partners to identify and solve real-life problems. Facilitators play an important role as guides to help students navigate project responsibilities and helping students in learning to learn, but they do not generally act as consultants or technical assistants. The EAST classroom mirrors the modern workplace by providing a dynamic environment in which students with all degrees of skills, experience, and aptitudes work together. Students are trained in and have access to advanced technical applications in architecture, animation, computer-aided drafting, database development, 3D design engineering, digital imagery, global positioning systems, geographical information systems, networking, system administration, programming, desktop publishing, digital filmmaking, and web development. These applications are provided not as a program objective but as tools to support student projects, the experience of which fosters studentsøteaming skills, responsibility, accountability, and personal initiative; mastering and applying basic skills and concepts; and creative and critical thinking skills.

The EAST classroom is characterized as existing in three parallel õenvironmentsöô the physical environment, which should be conducive to team work, accommodating of the use of technological resources, and enabling students to locate necessary materials; the learning environment, which entails guided instruction, class management that encourages responsibility,

flexibility and productivity, and projects which focus on self-directed, student-centered learning, community service, the use of advanced applications, and teamwork and peer mentoring; and an environment of expectations, which provides an open and encouraging culture in which students learn from their mistakes and that focuses on student development rather than content delivery, an outlook for program achievement that looks to the future instead of simply moving from day to day, and student work that is monitored to ensure that it is conducted productively so that students can solve problems, meet project goals, and use technology appropriately to meet those goals.

The EAST program also includes state-sponsored professional development geared towards providing orientation to district and school administrators, training for aspiring facilitators in instructional methods, program philosophies and expectations and technical systems administration, technology training for facilitators and students, and partnership conferences to disseminate and celebrate program achievements and conduct additional training workshops for students and facilitators.

Project Goals

The EAST evaluation involved two parallel and interdependent studies: a three year **implementation study** designed to provide insights into the variations in the implementation of the EAST program in different contexts throughout the state, and a two year **outcomes study** that sought to determine the impact of the initiative on participating facilitators and students by comparing eight schools who were randomly assigned from a larger pool of applicants to begin a new EAST program during the 2004-2005 school year, with a matched group of control students in eight schools that were assigned from the applicant pool as delayed implementation schools. Following are the major research questions that the EAST evaluation study sought to answer:

<u>Implementation Study</u>ô conducted in all existing EAST schools statewide throughout all three years of the project (2003-2006)

- What recruitment strategies are being used?
- What are the characteristics of EAST schools, students, and teachers?
- What are the nature, quality, and intensity of training activities for EAST facilitators and students?
- What are the nature, quality, and intensity of the EAST instructional practices and how do they differ in the various program settings?
- What are the nature and quality of the measures currently being used to document program implementation?
- What are the nature and quality of the assessments currently being used to measure teacher and student outcomes?
- What are the nature and quality of participating schoolsøpartnerships with local industry and community organizations, and what factors influence these partnerships at the local level?
- What are the nature, quality, and intensity of EAST parent involvement programs and activities?
- What are the nature and quality of administrative support for the program?

- In what ways and to what extent do the EAST National Center and the ADE support implementation of programs and practices in the participating schools?
- What are the nature and quality of dissemination activities offered to participating sites, and what is the level of participation in these activities?
- What steps have participating schools taken to sustain EAST beyond the three-year startup period?
- What are the costs associated with the EAST program?
- What obstacles, if any, have been encountered to program implementation, and what steps were or can be taken to overcome these obstacles?

<u>Outcomes Study</u>ô conducted during Years 2 and 3 of the evaluation (2004-2006) at 16 schools that were randomly assigned as new implementers (target) or delayed implementation (control) schools prior to the beginning of the study

- What is the impact of exposure to the EAST program on studentsø attitudes towards learning and school?
- What is the impact of exposure to the EAST program on studentsøproblem solving skills and content knowledge in math and English language arts?
- What is the initiative impact on facilitators attitudes towards teaching, classroom practices, and content knowledge?
- What influence do additional factors, including environmental factors, participant demographics, training conditions, instructional practices, and varying degrees of program integrity, appear to have on the impact of the EAST program on student and teacher outcomes?

Results from the Implementation Study

Data for the implementation study were obtained from facilitator and student focus group interviews, principal interviews, observations of EAST classes, and on-line principal and facilitator surveys that were administered in the spring of each year.

On the principal and facilitator surveys, schools reported using a combination of student recruitment and selection strategies that promoted general awareness of the program and those that targeted specific students. Since casting a wide net is not sufficient to ensure diversity and representativeness because of self-selection, the more targeted strategies would need to have been used to counterbalance such selection effects (by targeting recruitment towards the students who would be less likely to enroll on their own) if these goals were to be achieved. However, it was unclear specifically what kinds of criteria were used for the targeted strategies, and discrepancies between principalsø and facilitatorsø responses about what criteria were used as well as about who had primary responsibility for student placement imply that, to the extent that schools were making proactive efforts to ensure diversity, it is unclear how effective these efforts would have been. Comparisons of demographics of EAST students and non-EAST students in the same schools were conducted as part of the outcomes study and are discussed below.

Surveys also asked facilitators and principals to report on their perceptions of the impact of program training activities. Both groups have expressed very positive feelings about the training process throughout all three years of the study. When asked to rate the adequacy of EAST training for bringing their skills to the level they needed as a facilitator, facilitators reported that the training was best for raising their skills in the instructional methods advocated by EAST, and also provided favorable ratings to training they received in assessing their own progress in facilitating EAST. The lowest ratings were for training in the specific technology applications used in the lab, for which fewer than half of respondents felt the training was at least *adequate*. It is unclear, however, the extent to which these lower ratings reflected a problem with the training per se, or difficulty with the expectation of the EAST model that the facilitators should be able to guide students in the use of sophisticated technologies, even while having little or no expertise in those technologies themselves.

While ratings of training support were mostly very positive, ratings of other types of external program supports that were obtained from both principals and facilitators were more mixed. Availability of technology support, availability of funding to support participation in EAST training, opportunities for the facilitator to work with other EAST teachers, availability of support on logistics of program implementation, consistency of EAST philosophy with other school/district reforms, and time for facilitators to reflect on what they had learned were the highest rated. However, testing policies and practices, state-mandated curriculum frameworks, and grading policies and practices were seen as inhibiting effective implementation by substantial proportions of principals and facilitators.

Facilitators were more consistently positive about the support available from their school administrations, with large majorities agreeing that their principals supported class scheduling in their schools in ways that benefited the EAST program. Principals also echoed the facilitatorsø impressions about their willingness to provide flexibility and facilitate class scheduling. Principals also expressed strong agreement with almost all EAST philosophies, such as the value of learning in a real world context, encouraging self-directed learning, and the value of group work. The only notable exception was that a small majority of principals felt that *most students learn best in classes with students of similar abilities*. This was a somewhat surprising finding, given that the large majorities of principals saw EAST as integral to their schooløs overall academic program, although larger majorities saw the EAST program as benefiting *from* other coursework than the numbers who saw EAST as providing project opportunities or developing skills that were applicable *to* other courses.

Results from the Outcomes Study

EAST classroom observations were conducted in the eight study schools in winter 2004, spring 2005, fall 2005, and spring 2006. Each facilitator was observed a total of eight times, four times in each year. Additional measures that were used to assess the programøs impact on facilitatorsø fidelity to the model included the end-of-year program ratings (completed by consensus by the team of EAST, Inc. personnel at the end of each program year) and online principal and facilitator surveys. Since the end-of-year ratings aligned, for the most part, with the classroom observations, ratings from the latter instrument were used as the focus for analyses. In both years of the outcomes study, observations were consistently strong across almost all schools for physical environment; however, overall fidelity, and fidelity in educational

environment and in environment of expectationsô as well as in the six sub-ratings within these two environmentsô were much more variable and in both years reflected programs that ranged from those that were still performing near the low end of the scale to those that were already approaching model programs.

An important additional aspect of program fidelity that could not be addressed through classroom observations is the approach that EAST schools take to advertising and recruiting students for participation, and screening and placing interested students into EAST classes. It should be noted that the EAST model itself does not advocate that the program is appropriate for everyone, so a degree of prescreening may be appropriate; but if the study schools did not succeed in creating classes that were representative of the school population within the limitations specified by the model, it might not be possible to determine whether any study findings are applicable to types of students for whom the program *is* intended but who may have been underrepresented in the sample. Comments from school-based EAST staff in interviews and on surveys implied that there are categories of students (e.g., students with serious behavioral or attitude problems, õgrade seekers,ö and students who are accustomed to greater structure and more formal measures of progress) for whom some schools considered EAST to be less appropriate. As in the statewide surveys for the implementation study, surveys of the outcome-study schoolsørecruitment and screening procedures revealed a mix of strategies aimed at broad dissemination and those that were targeted towards specific students, but the criteria used for this targeting were not always clear, and the effectiveness with which criteria may have been implemented was called into question by the fact that most facilitators did not agree with their principalsø assertion that the principal played a role in recruitment and selection. Demographic comparisons revealed that EAST students were not representative of their nonparticipating peers on several demographic variables, including race, gender, grade level, Title I eligibility, gifted status, and special education status.

Other potentially important characteristics on which the groups might have differed but that were not measured included traits such as comfort with technology, self-discipline, ability to focus on long-term goals, and a willingness to take on responsibility, which the EAST developers consider necessary for EAST students to possess to at least some degree (and on which the EAST students would therefore not necessarily be expected to be representative of the school). Traits on which the developers thought that EAST students should be representative of the school included the ability to work independently, leadership ability, and willingness to work hard (among others), as well as hindering characteristics such as disciplinary problems, not taking school seriously, and emotional or socialization problems. On a Year 3 recruitment survey, very few schools reported that they made explicit efforts to obtain diversity on most of the latter characteristics. For this reason, and since self-selection would be likely to favor students who possess the positive traits and to weed out students who possess the hindering traits, it can not be assumed that EAST classes were representative of their schools on these The EAST director has indicated that it may not be preferable, from a characteristics. programmatic perspective, to make such recruitment issues a top priority while a program is first becoming established in its early years. Nevertheless, study results must be interpreted with this caveat in mind.

Analyses were conducted to estimate the impact of EAST participation on 16 outcomes in the areas of academic skills, problem solving skills, motivation for school, self-directed learning style, and motivation to pursue further learning. (There were differences in test and survey response rates by gender, grade level, and socioeconomic status, and among control students, by ethnicity and English proficiency; however, the resulting samples of EAST and control students with available outcome data remained statistically indistinguishable on most characteristics, and differences that did exist were modest. Conclusions about program impact should therefore be valid for the types of students participating in EAST at the study schools.)

Among the 16 student outcomes that were studied, analyses indicated that participation in EAST appears to have a positive, statistically reliable impact in five domains. These included three problem solving domains (defining the characteristics of a problem, assessing the outcomes of a solution, and revising strategies in response to the assessment of outcomes), one motivation domain (motivation for school derived from accomplishment), and self-directed learning style. The preponderance of evidence for program effects in the area of problem solving skills seems consistent with one of the most central goals of EAST, and may point to a particular strength of the program. Although no direct effects were found indicating an impact of the EAST program on studentsø math and reading test scores, this is a notoriously difficult relationship to demonstrate. Given the myriad of other factors that influence academic achievement and the limitations of standardized testing for measuring such skills, this should be taken as a failure to find a relationship, but certainly not as evidence that none exists. However, the domains on which EAST *has* been shown to have an impact are widely recognized as being important for both academic and career success.

There were a number of other factors that were found to reduce the likelihood of success in achieving some of these objectives, particularly coming from a low income family and attending an urban school, while the absence of initial skills and proclivities in many of the outcome areas (including motivation) would also act as a handicap. Importantly, however, findings indicated that while these conditions would put a student at a disadvantage, for the most part they were *not* observed to reduce the magnitude of the impact of EAST participation. In other wordsô in the case of motivation, for exampleô EAST can still be beneficial *in spite* of low initial motivation. The one notable exception where EAST was found to have provided less benefit for certain students was that, while participation was found to have an impact on developing studentsøproblem evaluation skills, it only proved to be beneficial in this domain for students who were initially unskilled at alternative generation. This should not be taken to mean, however, that students who *are* already skilled at alternative generation can not benefit from the program, since they could still benefit (at least) in the other four domains.

The fact that there were almost no indications of measured characteristics which reduced the observed impacts of EAST further underscores the importance of involving a diversity of students in the program. However, the ambiguities about the recruitment and selection procedures imply that schools may not be achieving this in all cases. If that is so, the programô at least in these newly implementing schoolsô might not be reaching certain groups of students who could benefit from the program as regularly as it might, especially those types of students who are less likely to seek it out. At the same time, if there are categories of students who tend not to enroll in EAST, it raises the possibility that the program impacts for such students might be smaller than the impacts observed in this study. In short, if enrollment is not representative of the schools, it could mean that some students who could benefit may be missing out, but could also mean that we cannot be sure whether there might be other students who might *not* benefit. While it may be inappropriate for ADE to dictate to the schools how they should conduct their recruitment efforts, it is important to be aware of the implications of these procedures for evaluating the program.

Recommendations

The following recommendations have emerged from this study:

Recommendations for the program.

- The successes observed in this study provide additional reason to continue to expand the program. However, because participating schools were self-selected, it should not be inferred that the program should be encouraged for any school who does not really want it.
- While the state may not want to dictate program recruitment and selection policies, the importance of more proactively reaching out to populations who are less likely to seek out the program, especially traditionally underserved or under-represented groups, might be further stressed as programs mature, with an emphasis that casting a wide net is not sufficient to fully ensure diversity.
- The observation rating form could serve as a useful tool for focusing discussions around õsite healthö visits.
- Further study of program impacts could be valuable for helping to identify whether the skills and attitude changes that EAST develops have long-term impacts on studentsø further education, or even on choice of career paths.
- Further study may also be warranted in order to more closely examine the extent to which program benefits may be influenced by other affective student characteristics, and the extent to which the program may be able to help students overcome certain characteristics such as a lack of interest in school. This could help further strengthen the value of the program by achieving a more targeted diversity, identifying and encouraging participation from types of students who might be shown to benefit but who traditionally might not apply. Since there are likely to be a number of student characteristics that are relevant to success other than the demographic indicators available from school records, school personnel who know the students may be in the best position to identify appropriate candidates. The targeted recruitment efforts that were often used by the schools can therefore be an effective way to accomplish this balance, as long as referring staff use appropriate criteria. We agree that diversity is a laudable goal for EAST enrollments, but it is not clear whether this should necessarily mean that enrollment should be representative of the school in all cases.
- Another potential objective for further research would be to explore the impacts of better established programs. It is quite possible that programs that have had the opportunity to become fully established might demonstrate even stronger impacts for a larger number of outcomes.

Recommendations for replication of the evaluation methodology.

- The value of development of a logic model early in the study can not be overemphasized. Particularly for a program such as EAST, for which adaptability and responsiveness to specific circumstances are among its most important distinguishing features, it is critical to clearly define the parameters within which this flexibility is meant to exist, both to help new practitioners understand what is expected of them, and to help researchers recognize the program that is being evaluated. This process can be quite time consuming, particularly if the program design has never been explicitly laid out before, but it is in this very situation when it is needed the most. It should be noted, however, that it may not be possible to create a written program description that is detailed enough for someone with no prior familiarity to implement or to evaluate the program. It may very well be that a program such as EAST requires personal training from existing staff to be replicated or evaluated, which would have significant implications for the cost of replication.
- Measuring program fidelity is a crucial part of program evaluation, especially when the programs that are being studied are in the earlier stages of implementation. Observation is an important method for collecting evidence of fidelity, especially when interactions between students and teachers form an essential part of the model. However, there may also be critical aspects of program fidelity that cannot be assessed through passive observation alone, especially if resources do not allow for frequent enough observations to obtain a representative picture of the program in a short period of time. In such cases, interviews and/or surveys of site-based program staff and of monitoring staff can provide an invaluable complement to observation, and can also access potentially important factors (such as teachersø attitudes and beliefs) that would be extremely difficult if not impossible to infer from observation.
- Assessing the possibility of the existence of selection effects can be exceedingly difficult. It may be worth implementing additional behavioral or attitudinal measures to try to evaluate student characteristics that are expected to influence program outcomes, even if they do not represent goals of the program per se. However, since it is presumably impossible to directly assess all potentially salient variables, it is equally important to make inferences about the likelihood of unmeasured selection effects by examining the procedures used to identify participants.
- Conducting a randomized evaluation design in a public school setting is extremely challenging, and in some circumstances, its pitfalls can be so severe that it may not be the best approach. Where the conditions exist that make it possibleô most notably, the opportunity to randomly select schools (or classrooms) for participation, or at least to randomly assign applicants to participating and control conditionsô there are a number of additional conditions that must be achieved in order to maintain the integrity of the study. One of the most important issues is to ensure the cooperation of participating schoolsô including the control schoolsô throughout the study. Possible strategies can include providing incentives, involving the schools in discussions about the value of the evaluation and in the process of designing the evaluation, providing schools with formative feedback throughout the study to make it more valuable, sanctioning of the evaluation from authorities, and laying out the obligations of participation in the evaluation as a condition for program participation. Nevertheless, it is important to avoid making schools feel like the study is being forced down their throats, as lack of

cooperation can invalidate the randomization process. This is most challenging when control schools are being asked to participate actively in the study. Unless the program is already considered desirable and can be offered as a reward for cooperation, it might be best to design the study in such a way that only the most minimal involvement is required from control schools.

Table of Contents

Table of Contents	X
List of Tables	, xiii
Introduction	1
Research Context	1
Purpose of the Research	3
Model to be Evaluated	3
Major Research Questions	9
Years 1-3 Implementation Study	11
Methods	11
Participants	11
Selection of Participating Schools	11
Selection of Participating Students	12
Measures/Instruments	13
Procedures	18
Results	18
Student Recruitment and Selection Strategies	19
Training and Orientation Activities	21
Program Support and Obstacles to Implementation	23
Year 2-3 Outcomes Study	25
Methods	25
Participants	25
Selection of Experimental and Control Schools	25
Selection of Control Students	29
Measures/Instruments	32
Logic Model	32
Classroom Observations	33
End-of-Year Program Ratings	45
Student Survey	46
Supplemental Recruitment Survey	47
Student Inventories	49
Iowa Tests	55
Facilitator and Student Focus Group Interviews	56
Procedures	56
Results	61
Fidelity of Program Implementation	61
EAST Classroom Observations	61
End-of-Year Program Ratings	69
Recruitment and screening procedures	71
Student Outcomes	88
Implications of Student Mobility and Instrument Response Rates	89
Impact of EAST on Student Skills and Content Knowledge	92
Conclusions	95
Limitations	98

Recommendations	99
Sustainability, Capacity Building and Dissemination	102
Bibliography	104
Appendices	107
Appendix I: EAST Logic Map and Detailed Logic Model	A-1
Appendix II: Facilitator and Student Focus Group Interview Protocols	A-14
Appendix III: Spring 2004 Administrator Interview Protocol	A-32
Appendix IV: Year 1 Classroom Observation Protocol & Procedure	A-35
Appendix V: On-line Principal Surveys	A-40
Appendix VI: On-line Facilitator Surveys	A-73
Appendix VII: Principal and Facilitator Surveys of Non-Study Schools:	
Student Recruitment and Selection	.A-114
Appendix VIII: Principal and Facilitator Surveys of Non-Study Schools:	
Staff Training Activities	.A-120
Appendix IX: Principal and Facilitator Surveys of Non-Study Schools:	
Program Support and Obstacles to Implementation	.A-125
Appendix X: Principal and Facilitator Surveys of Non-Study Schools: Additional Data	.A-137
Appendix XI: Statement of Commitments	.A-146
Appendix XII: Results of Control Group Matching for 2004-2005 School Year	.A-152
Appendix XIII: Demographic Comparisons of Target EAST Students	
and Control Schools	.A-155
Appendix XIV: Outcomes Study Classroom Observation Protocol,	
Classroom Observation Annotated Guide, Fidelity Observations Guidelines	.A-159
Appendix XV: Pearson Correlations Among Fidelity Observation Subscales	
Fall 2005 and Spring 2006 Observations	.A-188
Appendix XVI: End-of-Year Program Fidelity Ratings Protocol and Guidelines	.A-189
Appendix XVIIa: EAST Student Survey	.A-200
Appendix XVIIb: Control Student Survey	.A-205
Appendix XVIII: 2006 Supplemental Recruitment Survey	.A-210
Appendix XIXa: Fall 2005 Student Inventories ó	
Inventory of School Motivation	
Confirmatory Factor Analyses	.A-220
Appendix XIXb: Fall 2005 Student Inventories ó	
Social Problem Solving Inventory for Adolescents	
Confirmatory Factor Analyses	.A-226
Appendix XX: Demographic Comparisons	
Target EAST Students vs. non-EAST Students in EAST Study Schools	.A-232
Appendix XXI: Demographic Comparisons of Leavers and Stayers of EAST Students	.A-240
Appendix XXII:Demographic Comparisons of Respondents vs.	
Non-Respondents & EAST Students	.A-244
Appendix XXIII: Demographic Comparisons of Respondents and	1 252
Non-Respondents o Control Students	.A-253
Appendix AAIV: Demographic Comparisons of EASI and Control Students	1 262
among 11 b5 Kespondents	.A-202
Appendix AAV: Demographic Comparisons of EAST and Control Students	1 266
Among student inventory respondents	.A-200

Appendix XXVI: Results of Hierarchical Linear Modeling of Predictors	
of Student Outcomes	70

List of Tables

Fable 1 Demographics of Target Schools for Year 1 Implementation Study 12
Table 2 Overview of EAST Statewide Implementation Study
Fable 3 Professional Development Activities for EAST Facilitators
Fable 4 Demographics of Target Schools for Year 2-3 Outcomes Study
Fable 5 Control Group Matching Criteria ó Cohort 1
Fable 6 Control Group Matching Criteria ó Cohort 2
Fable 7 January 2005 EAST Pilot Observations Interrater Reliability Based on Percent
Agreement
Fable 8 October 2005 EAST Observations
Interrater Reliability Based on Percent Agreement
Fable 9 January 2005 EAST Pilot Observations
Interrater Reliability Based on Cohenøs Kappa
Fable 10 October 2005 EAST Pilot Observations
Interrater Reliability Based on Cohenøs Kappa
Fable 11a January 2005 Pilot Observations
Observer Consistency in % of õDon¢t Knowö and Missing Ratings Physical Environment41
Fable 11b January 2005 Pilot Observations Observer Consistency
in % of õDon¢ Knowö and Missing Ratings Educational Environment
Fable 11c January 2005 Pilot Observations Observer Consistency
in % of õDon¢ Knowö and Missing Ratings Environment of Expectations
Fable 11d January 2005 Pilot Observations Observer Consistency
in % of õDon¢ Knowö and Missing Ratings Overall Fidelity
Fable 12a October 2005 Group Observations Observer Consistency
in % of õDon¢t Knowö and Missing Ratings Physical Environment
Fable 12b October 2005 Group Observations Observer Consistency
in % of õDon¢t Knowö and Missing Ratings Educational Environment
Fable 12c October 2005 Group Observations Observer Consistency
in % of õDon¢t Knowö and Missing Ratings Environment of Expectations
Fable 12d October 2005 Group Observations Observer Consistency
in % of õDon¢t Knowö and Missing Ratings Overall Fidelity
Fable 13 Summary of Characteristics of the Original ISM and SPSI-A Instruments
Fable 14 Student Inventoryô ISM Scales Internal Reliability Analyses
Fable 15 Student Inventoryô SPSI-A Scales Internal Reliability Analyses
Fable 17 EAST Outcome Study: Outcome Variables and How They Were Measured
Fable 18a EAST Outcome Study: Student Level Explanatory Variables
Fable 18b EAST Outcome Study: School and Program Level Explanatory Variables
Fable 19 SY 2004-2005 Classroom Observation Forms
Characteristics of Participating Classes and Students
Table 20 SY 2004-2005 Classroom Observation Forms Summary of Ratings & All Classes63
Fable 21 SY 2004-2005 Classroom Observation Forms
Summary of Ratings by Program Level
Fable 22 SY 2004-2005 Classroom Observation Forms Summary of Ratings by Locale
Fable 23 SY 2005-2006 Classroom Observation Forms
Characteristics of Participating Students

Table 24 SY 2005-2006 Classroom Observation Forms Summary of Ratings ó All Classes67 Table 25 SY 2005-2006 Classroom Observation Forms
Summary of Ratings by Program Level
Table 26 SY 2005-2006 Classroom Observation Forms Summary of Ratings by Locale 69
Table 27 SY 2004-2005 End-of-Year Program Rating Forms
Summary of Ratings ó All Schools70
Table 28 SY 2005-2006 End-of-Year Program Rating Forms
Summary of Ratings ó All Schools70
Table 29 Principal and Facilitator Surveys Student Recruitment Strategies
Table 30 Principal and Facilitator Surveys Student Selection Procedures 73
Table 31 Principal and Facilitator Surveys
Agreement Between Principals and Facilitators About Recruitment Strategies74
Table 32 Principal and Facilitator Surveys
Agreement Between Principals and Facilitators About Selection Procedures75
Table 33 Spring 2006 Surveys Number of schools with
Agreement Between Principals and Facilitators About Recruitment Responsibilities77
Table 34 Spring 2006 Surveys Number of Schools with Agreement Between
Principals and Facilitators About Selection/Placement Responsibilities
Table 35 Spring 2006 Surveys Agreement Between Principals and Facilitators
About Final Decision About Placement
Table 36a Spring 2006 Recruitment and Screening Surveys
Potentially Helpful Student Characteristics
Table 36b Spring 2006 Recruitment and Screening Surveys
Potentially Hindering Student Characteristics
Table 36c Spring 2006 Recruitment and Screening Surveys Demographic Characteristics 87
Table 37 Response Rates Among EAST and Control Students 91
Table 38 Domains on Which Relationship to EAST Participation Was Found

Research Context

In his 2001 address to the Department of Labor, Federal Reserve Board Chairman Alan Greenspan reflected that the õenormous new capacity [of technology] to capture, analyze and disseminate information has begun to alter significantly how we do business and create economic value,ö with the consequence that our countryøs market-driven educational system must equip students only with technical know-how, but with the ability to create, analyze and transform information and to effectively interact electronically with others.ö Indeed, more than two decades earlier, the seminal report õA Nation at Riskö noted, õThe people of the United States need to know that individuals in our society who do not possess the levels of skill, literacy and training essential to this new era will be effectively disenfranchised, not simply from the material rewards that accompany competent performance, but also from the chance to participate fully in our national lifeö (National Commission on Excellence in Education, 1983). The emergence of this onew economyo signals a future society that differs greatly from the one in which past generations have grown, worked, and succeeded. The metamorphosis of our society has been and will continue to take us from being an industrialized nation to that of a cohort of õknowledge workersö (Druker, 1994). The transition of the U.S. from an industrial to an information society clearly implies that in education, teaching reading and writing is no longer sufficient, although these are certainly the foundation skills upon which other literacies (including technological, visual, and numerical) are built (Murray, 2003). This group of workers will require a unique set of primary skills that will allow them to flourish and increase productivity. The economy will need new types of workers with important thinking, deductive reasoning, and creativity skills who are well versed in the õbasicsö of education, as well as in critical and analytical thinking and problem solving. In his book, The New Basics (2002), David Thornburg wrote that there is a need in society for creativity, technological fluency, problemsolving ability, and entrepreneurial skills that are not being taught in our schools (Thornburg, 2005). Problem solving, self-directed learning, and motivation are all areas that are of clear importance to the creation of a cohort of knowledge workers. In his seminal work, Taxonomy of Educational Objectives, Bloom (1984) identified a hierarchy of skills important to learning, including knowledge, comprehension, application, analysis, synthesis, and evaluation. Too often classrooms focus much of their time on the first three types of learning, in large part due to the increased focus on student academic achievement based on standardized tests. For students to meet the needs of tomorrow knowledge society, it is essential that teachers focus more on the higher level skillsô analysis, synthesis, and evaluation, a unique blend of skills essential for the future of our country. Indeed, the skills necessary for success in todayøs world differ so greatly from those of the past that many of tomorrow jobs do not yet exist.

The classrooms of today and those being developed for tomorrow provide a unique opportunity to assist in the preparation of students for the future economy. U.S. Education Secretary Margaret Spellings said of technology in the classroom that it is õí changing how students learn. We can engage students in new ways and transcend the walls of the traditional classroomö (CDW-G, 2005). Her comment is particularly meaningful in light of the need for focus in todayøs learning environments on allowing students to become problem solvers, motivated learners, proficient technology users, and critical and analytical thinkers. EAST

program staff believe that the use of technology and different learning environments are significant ways to help students reach this potential. The hope is that by providing solid technological skills, students will be able to communicate, link with one another, and create learning environments that are strongly collaborative and mutually beneficial. While it would not be possible to create an EAST-like environment in every classroom in the world, by providing these tools to students (as they are provided to EAST students) it may be possible to foster a group of global learners who embody the skill set identified as essential, including (as mentioned above) problem solving, motivation, and self-directed learning. If students are to lead us in a positive direction, then it is vital that they have access to the tools and thought processes necessary for success in the future world.

Evaluation of student learning outcomes and of technology in general is of paramount importance to validating the work of programs and organizations intent upon changing the structure of todays classrooms. As information technology continues to become standard equipment in the classroom, educators need insight into how to maximize its positive impact (Ringstaff & Kelley, 2002). One means of providing educators with such lessons is through careful evaluation of current programs (Johnston & Barker, 2002). In 1998, the authors of the Milken report, Technology in American Schools: Seven Dimensions for Gauging Progress, noted that while accountability for public investment in education technology was often reported in terms of quantity of equipment and connectivity, policymakers needed to examine the qualitative impact of technology, yet appropriate measures for doing so had not yet been developed. The report of The Secretary's Conference on Educational Technology (McNabb, Hawkes, & Rouk, 1999) echoed this problem by noting that õif resources are to be expended on technology, it is becoming a political, economic, and public policy necessity to demonstrate its vital effectivenessö and concluded with a call for rigorous, innovative, and relevant technology evaluation designs. One outcome of the conference was an acknowledgement that while policymakers tend to value summative reports that document student achievement, teachers and administrators value formative reports that document implementation outcomes that help them make informed decisions about technology programs. A study of the issues surrounding technology education assessment and evaluation (Reeves, 1998) found that õtraditional assessment measures and simplistic evaluation designs are unlikely to reveal the complexities of the nature and outcomes of programs that differ from the traditional paradigm of instruction.ö Indeed, it is essential that evaluations of technology and learning programs explore all aspects of implementation and outcomes, including those õinitialö outcomes of student learning such as problem solving and self-directed learning skills that are essential to success in the information age.

Since 1998, numerous evaluations of technology education programs have been conducted. A review of these in *The Learning Return on Our Educational Technology Investment* (Ringstaff & Kelley, 2002) concluded with the statement that õmeasuring the impact of technology use on student achievement is fraught with difficulties.í Classrooms are not experimental laboratories where scientists can compare the effectiveness of technology-based classrooms to traditional instructional methods while holding all other variables constant.í Few reliable, valid, and cost-effective assessments exist that measure studentsøhigher-order thinking skills, problem-solving ability, or capacity to locate, evaluate, and use information.ö Ringstaff and Kelly suggest that rather than examining the cost-effectiveness of technology education

programs, researchers examine the conditions under which technology offers students the most benefits.

This evaluation of the EAST initiative is aligned with and supportive of the U.S. Department of Education¢ goals as described in the *No Child Left Behind Act* (NCLB) and the U.S. Department of Education¢ 2002-2007 Strategic Plan. The evaluation assessed a middle and high school technology education project that promotes equity of educational opportunities and works to raise student achievement through a focus on higher-order thinking skills and innovative uses of educational technology. Moreover, the project includes an extensive professional development component and thus supports the NCLB goal of preparing and training high-quality teachers. Information gathered through the proposed project has been disaggregated according to NCLB requirements in an effort to provide evidence of the impact of the project on students of diverse backgrounds and abilities.

Purpose of the Research

Model to be Evaluated

The Environmental and Spatial Technology (EAST) Initiative is a performance-based learning environment utilizing project-based service learning, integrated with advanced technological applications in an interdisciplinary environment where the intellectual and problem-solving growth of students is the focus. The project, which began in 1995 with 20 students in an Arkansas high school, has since expanded to over 190 schools in eight states (Arkansas, California, Hawaii, Illinois, Iowa, Louisiana, Mississippi and Pennsylvania). At the outset of the current study (school year 2003-2004), Arkansas had approximately 130 EAST programs. The majority of EAST sites in Arkansas were in high schools (90%), with seven percent at middle schools, and another three EAST programs that were implemented in an elementary school or college. Sites are spread over the predominantly rural state of Arkansas, with the majority in rural counties, and the remainder divided roughly evenly between suburban and urban counties.

EAST¢s central concept of the importance of students¢ responsibility for their own learning, with a focus on problem solving, is validated by a recent study that found that the most successful technology applications were those that required students to think for themselves and apply technology to complex, extended tasks (Means, Penuel, & Padilla, 2001). Moreover, research shows that cooperative learning represents a valuable strategy for helping students attain high academic standards (Kagan, 1993; Cohen, 1994). After nearly 50 years of research, there is strong agreement that cooperative methods can and usually do have positive effects on students¢ achievement, particularly when they include interdependence and individual accountability (Holt, 1997), as in the EAST model. In addition, community service programs teach students higher-order cognitive processes and skills, such as data gathering, critical thinking, and decision-making (Kirby, 1989). It is believed that encouraging EAST students to show initiative and become engaged in their learning process develops them into active learners whose high degree of responsibility ensures the success of team efforts.

Student Selection and Participation

EAST staff work with the local schools to recruit a diverse group of students interested in taking the EAST elective, with diversity understood to include aptitude, academic motivation, and achievement levels, as well as race, gender, and economic background. Students with special needs, gifted students, at-risk students, and all students in between are placed on a level playing field by being in the same class. Ultimately, the goal should be to construct classes that are representative of the demographics of the school as a whole.

In an EAST class, students mentor one another and develop personal accountability and a true appreciation of diversity, while acquiring ownership in both the project and the learning process as a whole. Typically, between 15 and 20 students enroll in a single section of EAST. The number of sections offered per site is typically dependant upon the size of the school, with larger urban and suburban schools enrolling upwards of 100 students per credit cycle and smaller, rural schools enrolling 40-50 students per credit cycle. Classes meet in step with the school schedule (daily for schools that offer six- or seven-period days or four-by-four block schedules, and every other day for schools on A/B block scheduling), while students may also spend additional time outside the regular class schedule and/or outside the classroom working on their projects as needed.

Instructional Model

Teachers are chosen by their schools and in turn are trained by EAST staff to become õfacilitatorsö who guide and assist students as they pursue their projects. Facilitators also act as resources for the projects and as managers who direct all the projects being undertaken at the site.

The EAST classroom mirrors the modern workplace by providing a dynamic environment in which students with all degrees of skills, experience, and aptitudes work together. Students are fully engaged in, and responsible for, their own learning. The EAST classroom, a physical space with associated equipment specifically configured to EAST standards and purchased by the school, usually with support from the Arkansas Department of Education (ADE), is the place where EAST is primarily experienced. All student activities do not take place in the EAST classroom, however. Often, students are required to be in the field (in their community or various areas of the school).

The EAST model includes a multifaceted, interdisciplinary, service-learning curriculum integrated with advanced technical applications from the following disciplines: architecture, animation and visualization, computer aided drafting (CAD), database development, 3D design dngineering, digital imagery, global positioning systems (GPS), geographical information systems (GIS), networking, system administration, programming, desktop publishing, digital filmmaking, and web development. However, the teaching of these applications is secondary to a focus on allowing students to develop as learners in an integrated setting that fosters such diverse skills as teaming, responsibility, accountability, and personal initiative. While students concentrate on projects, their facilitator focuses on developing studentsø intellectual capacities through the application of traditional curricular objectives. Student success is measured by the

degree of growth the students have demonstrated in mastering and applying basic skills and concepts, as well as progress made toward the development of creative and critical thinking skills.

To develop these skills, students connect with peers, faculty, community members, and business and university partners, who help them to identify needs in the community. Students work directly with these partners to develop strategies to help address these needs, utilizing the partnersøexpertise and the technological resources available through EAST. From the studentsø point of view, project-based service learning does not resemble traditional school activities. Rather, it resembles õworkö: students are expected to produce real, tangible results of high quality. Students involved in the EAST Initiative have the opportunity to identify and solve real-life problems instead of participating in simulations and/or õwhat ifö situations. Often they leave a legacy of starting or completing a project that benefits their community well beyond the time of their graduation.

EAST facilitators play an important role as guides to help students navigate these responsibilities. However, their role is as important for what the facilitators do *not* do as for what they do. The facilitator is responsible for establishing a classroom environment that supports and guides studentsøproject work and personal development, but she does *not* generally act as a consultant or technical assistant. Indeed, between the student teams and their project partners, the facilitator is often the least knowledgeable about the specifics of the problem or the technological tools being applied. It is not the facilitatorøs responsibility to be the expert on every project; rather, it is to guide students in learning to learn, to help them recognize the need to develop organizational and planning skills, and to help them learn to recognize the implications of their decisions. The philosophy of the EAST model is that students are better able to learn these concepts through direct experience rather than having them õtaught.ö

The nature of the EAST classroom which facilitators are responsible for creating is characterized as existing in three parallel õenvironmentsöô the physical environment, the educational environment, and the environment of expectations. The modeløs definition of how these environments are constituted (and the facilitatorsø responsibilities in establishing them) is summarized below; this model is described in greater detail in the logic model in Appendix I.

• Physical Environment

The facilitator creates a classroom with a physical layout that is conducive to team work, accommodating of the placement and use of technological resources, and which enables students to locate necessary materials.

• Educational Environment

Instruction in the EAST classroom is characterized by a facilitator who models effective learning and development of real-world skills, and points students in the right direction without giving answers directly. The facilitator adjusts his or her level of support as appropriate to the needs of specific students.

Class/Project Management in EAST is defined in terms of goals and responsibilities as well as flexibility to meet project needs. The facilitator provides guidance and direction to meet goals and helps project teams to capitalize on the diversity of students in terms of aptitude, experience, knowledge, interests, and learning styles. An effective EAST facilitator expects productive work from students and addresses inappropriate behavior or lack of productivity in different ways depending on individual student needs. Participation and group work are encouraged from everyone. The facilitator helps students obtain resources and establish partnerships and serves as a role model and a liaison between students and the outside world.

The Nature of Projects in an EAST classroom includes a focus on the four pillars of EAST learning, which include self-directed, student-centered learning; community service; the use of advanced technological applications; and teamwork and peer mentoring. Students are encouraged to select their own projects, which should be practical in nature and encourage learning that extends the learnerøs view and understanding and/or skills, and which should provide opportunities for students to practice planning, executing, and judging the status and quality of the project and project goals.

• Environment of Expectations

EAST Classroom Culture is positive, welcoming, and encouraging; it involves a focus on the needs of students, on student development rather than content delivery, and on process over product, and there is an openness to intuitive thinking and the exploration of unlikely solutions. There is recognition that student growth takes place not only through success but also on the road to discovery. Students are given the freedom to think through ideas and evaluate them, while the facilitator provides a safety net by assuring students that it is safe to fail or make mistakes.

Outlook for Overall Program Achievement should be reflected by an EAST facilitator who looks to the future instead of simply moving from day to day. The facilitator takes a long view of program success, looking to the strengths of the community and students and placing value on student growth over time.

Student Work is expected to be conducted productively, typically in cooperative groups, with all students working daily on projects that interest them. The work of students is monitored in a clear and organized manner by their peers as well as by the facilitator, and students are expected to utilize time effectively, solve problems, meet project goals, and use technology appropriately to meet those goals.

Professional Development

Participants in school reform must have training before implementation, for one of the key factors in sustaining education reform is expanding educatorsø knowledge base (Klein, 1996). Furthermore, because group work dramatically changes the teacherøs role, professional development is vital to the implementation of cooperative learning (Cohen, 1994). Successful implementation of the EAST model requires professional development for facilitators and administrators to support facilitatorsøroles as facilitators of learning. Initial training focuses on facilitatorsø participation in teams that model the approach to be used in the classroom. As the

facilitators work together to address their practice õproblemö during the training session, they also discuss ways of solving pedagogical problems associated with team learning. As part of their training, facilitators assess their own teaching styles and develop plans for implementing project-based learning.

EAST¢s extensive professional development amounts to 24 days of required training in the first year (with additional opportunities encouraged) and six days of training in each of the second and third years. Training for new facilitators is conducted in three distinct phases, as described below.

In Phase I, aspiring facilitators are given hands-on instruction in classroom and systems management using the same equipment and systems they will use in their own classes, before their schooløs EAST program begins operation. Workshop participants meet experienced EAST facilitators and students as mentors and gain insight into the creation of successful programs. They are taught how to stay focused on the intellectual development of their students (given individual learning styles), how to properly evaluate student progress, and how to set high expectations of their students. Because the key to success relies heavily on the educational philosophy of the individual facilitator, visiting facilitators are exposed to strategies that help move students from the traditional student-centered/teacher-directed educational environment to a collaborative work environment. Upon completion of this phase of professional development, participants present their implementation plan for review by EAST staff, thus ensuring a positive beginning for their schooløs EAST implementation.

Phase II includes a two-day workshop and discussion sessions that help facilitators sharpen their understanding of the role of the facilitator as they begin to guide students in developing school and community service projects. The workshops provide opportunities for facilitators to increase their knowledge of the application options available through EAST and to share their successes and difficulties. Facilitators spend substantial amounts of time in Phase II working in groups to learn from each other. Additional subjects include problem-solving strategies and reinforcement of Phase I professional development, with more detailed presentations given to support the facilitators in the areas of equity in the EAST classroom; marketing the local EAST program; and recruiting for the following semester or school year.

Phase III consists of a two-day workshop designed to make facilitators who attended the Phase II session more knowledgeable about the corporate culture of knowledge-based companies; representatives from EAST business partners and civic leaders often attend. Attendees discuss how to develop partnerships in the public and private sectors, and how to fully involve those partners in facilitating student development within the EAST class.

<u>Technical Professional Development.</u> An initial systems administration class is included as part of EAST implementation. This one- to three-day class usually takes place about 30 days after installation of the lab. A successful method of systems administration geared to the EAST environment has been developed as a result of lessons learned from developmental sites. Experienced EAST facilitators, along with Certified Network Administrators, develop curricula and help deliver specific training in systems administration during workshops for new facilitators. <u>Specific Technology Training</u>. Facilitators involved in the duplication of this model must grasp the conceptual capabilities of computer-aided design, design engineering, visualization, and geo-processing technology and techniques. Although they are *not* expected to become experts in these applicationsô that responsibility rests with the studentsô facilitators must have a basic understanding of the functionality, fundamental skills, and appropriate applications of todayøs relevant technological tools. In order to fully integrate advanced applications into the school curriculum, õseedö training must occur in the areas of geographic information systems (GIS), global positioning systems (GPS), architecture, engineering, and animation. Basic õbeginning pointsö of understanding advanced applications must be learned and transferred to new sites. The purpose of this training is to establish a base of understanding and awareness in areas of sophisticated software totally unavailable to most secondary educators.

Workshops for facilitators and students are offered throughout the year to establish a basic technical knowledge framework. Students and facilitators who are proficient in sharing what they learn attend Training Camps designed to teach specific basic skills in advanced applications. Upon completion of the training, students return to their respective schools to peer teach and facilitate other studentsø work using similar advanced applications. These camps are held near the end of the first semester, after EAST students have begun to develop an understanding of the basic concepts involved with the digital environment as well as entry-level applications. Facilitators are asked to select two students who will attend the camp with them.

Additional professional development opportunities include multimedia training tutorials that are classroom learning resources for students and facilitators, and facilitator õshadowingö opportunities for hands-on training in practical applications. Shadow training allows facilitators to experience the real needs of employers today, thus enabling them to lead their students toward the development of the skills necessary to ensure their future employment. Shadow training may be limited to those areas where public and/or private sector resources are available and willing to participate. Continuing education, an essential ingredient in the maintenance of relevant learning environments, is available to facilitators and students, who may access Internet sites specifically designed to inform users of advancements in technology. The annual three-day EAST Facilitator Summer Seminar attracts facilitators who have completed at least one year of EAST and provides a forum for facilitators to share best practices and address common needs.

The EAST Partnership Conference, an annual event most recently held at the Hot Springs, Arkansas, Convention Center,¹ celebrates the achievements of participants. This event, which in every sense except the age of the presenters has the look and feel of a professional conference, includes training sessions in technical and pedagogical areas, a general exposition during which participants can discuss concerns, and opportunities for students to showcase their projects. During this exposition, a panel of volunteer judges selected by the EAST staff observe studentsø presentations of their projects and rate them on dimensions such as the projectøs value for community service and evidence that the project utilizes and develops self-directed learning, problem solving, communication and collaboration skills. Since 2005, in order to better emphasize the collaborative nature of the EAST model, schools have been rated for their overall

¹ Originally at the Little Rock Convention Center.

program rather than on individual projects. An awards ceremony is held towards the end of the conference to recognize the most successful programs for the year.

Community and Business Support

From its inception, EAST has promoted collaborations of business and institutions of higher education on projects involving students and their communities. High-tech companies have donated software to develop EAST classroom labs, and the University of Arkansasøs Center for Advanced Spatial Technology (CAST) has been providing training in the areas of GPS, GIS and architectural technologies. EAST currently has partnerships with more than 30 Arkansas businesses that provide schools with a broad array of technology.

Major Research Questions

Two parallel and complementary studies were conducted by Metis: an **implementation study** designed to provide insights into the variations in the implementation of the EAST initiative in different contexts throughout the state, and an **outcomes study** that sought to determine the impact of the initiative on participating facilitators and students.² These two studies were designed to be mutually supportive, each informing the development and direction of the other. For example, insights about program variations obtained through the first year of the implementation study were used to inform the development of instruments for the outcomes study, while preliminary findings from outcomes measures brought attention to additional questions that were further explored through the implementation study.

The implementation study, which was conducted in all existing EAST schools statewide throughout all three years of the project (2003-2006), was designed to answer research questions focused on developing an understanding of the process of creating and operating an EAST program in varying environments, including middle and high schools as well as rural and urban settings. The ultimate goal of this study was to gather, analyze, and triangulate data to create a clear picture of the EAST model of implementation in order to determine how it varies throughout the state, and to identify some of the supports and obstacles to implementation. Listed below are the research questions that the implementation study sought to address.

- What recruitment strategies are being used?
- What are the characteristics of EAST and non-EAST schools, students, and facilitators?
- What are the nature, quality, and intensity of training activities for EAST facilitators and students?
- What are the nature, quality, and intensity of the EAST instructional practices and how do they differ in the various program settings?
- What are the nature and quality of the measures currently being used to document program implementation? To what extent are these measures designed to assess the alignment of local practices with EASTøs educational principles?

² With the approval of the USDOE, a sustainability study that was initially planned was not fully implemented, due to the need to reallocate resources in order to ensure a rigorous experimental design for the outcomes study. Nevertheless, findings about sustainability were obtained and are discussed at the end of this report.

- What are the nature and quality of the assessments currently being used to measure facilitator and student outcomes? To what extent are these student performance assessments aligned with state content, industry, and technology standards?
- What are the nature and quality of participating schoolsøpartnerships with local industry and community organizations, and what factors influence these partnerships at the local level?
- What are the nature, quality, and intensity of EAST parent involvement programs and activities?
- What are the nature and quality of administrative support for the program?
- In what ways and to what extent do the EAST National Center and the ADE support implementation of programs and practices in the participating schools?
- What are the nature and quality of dissemination activities offered to participating sites, and what is the level of participation in these activities?
- What steps have participating schools taken to sustain EAST beyond the three-year startup period?
- What are the costs associated with the EAST program?
- What obstacles, if any, have been encountered to program implementation, and what steps were or can be taken to overcome these obstacles?

The outcomes study was conducted during Years 2 and 3 of the evaluation (2004-2006) at 16 schools that were randomly assigned as target or control schools prior to the beginning of the study.³ This portion of the evaluation study was designed to measure the impact of the EAST initiative on students and facilitators by attempting to answer the following research questions:

- What is the impact of exposure to the EAST learning environment and instructional practices on participating studentsø attitudes towards learning and school, including their aspirations to pursue postsecondary education, professional careers, and community service? How do these outcomes vary by sub-group?
- What is the impact of exposure to the EAST learning environment and instructional practices on studentsø skills (e.g., critical thinking, problem solving, communication, interpersonal, technology) and content knowledge in science, math, and English language arts? How do these outcomes vary by sub-group?
- What is the initiative impact on facilitators attitudes towards teaching, classroom practices, and content knowledge?
- What influence do additional factors, including environmental factors, participant demographics, training conditions, instructional practices, and varying degrees of program integrity, appear to have on the impact of the EAST program on student and facilitator outcomes?

³ Procedures for selection and assignment of participating schools and students are discussed in the Methods section of this report.

Methods

Following is a summary of each of the data collection instruments, procedures, and analysis methods that were utilized throughout the three-year study, including a description for each instrument of the administration schedule, sample selection procedure (where applicable), and statistical tests and/or qualitative analysis methods that have been applied to the data. In each case, the data collection instrument or procedure and analysis methods are also described in terms of how they relate to the projectøs research questions.

Participants

Selection of Participating Schools

The evaluator conducted site visits to six participating EAST schools in March 2004 that included observations of EAST classes and in-person interviews with each schooløs principal. The six target schools were selected to be representative of the majority of existing EAST schools in Arkansas (middle and high school programs), based on the following criteria:

- Target schools represented both middle and high schools or, in some cases, schools that served students at grades 7 through 12 but conducted an EAST program only at the high school level. Because the large majority of existing EAST programs were being implemented at the high school level, high schools were over represented in the sample (see Table 1 below).
- Target schools represented different types of geographic locales. Two locale categories were defined by collapsing the eight National Center for Education Statistics (NCES) Locale Codes for the 2002-2003 school year⁴ into two groups that represented an intuitively meaningful separation, based on ADE staff¢s familiarity with the communities represented by each code. The two categories that were developed were *large town/small city*, which included schools with NCES codes of 2 (mid-size city), 3 (urban fringe of a large city), 4 (urban fringe of a mid-size city) and 5 (large town); and *rural/small town*, which included schools with NCES codes of 6 (small town), 7 (rural, outside Metropolitan Statistical Area [MSA]) and 8 (rural, inside MSA).⁵ For brevity, these categories hereinafter will be referred to as *urban* and *rural*, respectively.
- Within each level (middle schools and high schools), target schools were also selected to include at least one high-poverty and one low-poverty school, at least one large and one small school, and at least one high-fidelity and one low-fidelity school. Socioeconomic status (SES) and school size groupings were defined by identifying quartile distributions of all EAST schools in Arkansas for the percentage of students eligible for free or reduced price lunch, and the total number of students, respectively. Where possible, schools were selected from the first and fourth quartiles of these distributions or, where this was not practical, from the lower end of the second quartiles and the upper end of the

⁴ See <u>http://nces.ed.gov/surveys/ruraled/index.asp</u>.

⁵ At the time these data were collected, no schools in Arkansas were located in communities with an NCES locale code of 1.

third quartiles. Fidelity was defined by the National Program Director, who was asked to estimate how closely existing EAST schools were implementing the national model and to recommend schools with a range of perceived levels of accordance (fidelity) with the model.⁶

- The sample was created to include at least two first-year EAST schools, including one that was considered high fidelity and at least one that was considered low fidelity.
- Finally, since the purpose of this phase of the study was descriptive rather than evaluative, the schools were not selected randomly but were chosen to be within a reasonable distance from Little Rock.

Six target schools were chosen within a 100-mile radius of Little Rock. These schools were distributed as indicated in the table below.

Demographics of farget behoods for fear fimplementation brady								
Locale	Middle Schools	High Schools						
Category								
	School 1: size = 4^{th} quartile, poverty = 4^{th}	School 3: size = 4^{th} quartile, poverty = 1^{st}						
	quartile, fourth year	quartile, fifth year						
Urban								
		School 4: size = 2^{nd} quartile, poverty = 4^{th}						
		quartile, first year						
	School 2: size = 1^{st} quartile, poverty = 4^{th}	School 5: size = 3^{rd} quartile, poverty = 1^{st}						
	quartile, first year	quartile, third year						
Rural								
		School 6: size = 4^{th} quartile, poverty = 1^{st}						
		quartile, fourth year						

 Table 1

 Demographics of Target Schools for Year 1 Implementation Study

During Years 2 and 3, activities for both the implementation study and the outcomes study were conducted at a separate group of schools that were selected for the outcomes study. Selection of these schools is described in the Participants section of the outcomes study below.

Selection of Participating Students

A sample of EAST facilitators and students participated in focus groups that were conducted during the EAST Partnership Conference during the winter of each year of the study. For the Year 1 interviews (February 2004), the evaluator asked the EAST National Program Director to identify a total of at least six but no more than ten facilitators from EAST schools (not necessarily the target schools) conducting programs for high school grades, and a similar number from schools with middle school programs. For each group, the Director was also asked to include participants from schools that represented a range of experience with the program and a range of perceived levels of accordance with the model, as described above. In addition, the Director was asked to identify facilitators whom he felt would be articulate and comfortable participating in group discussions. Each facilitator was then asked to identify one student from his or her school who met the same criteria of expressiveness and comfort. The middle and high

⁶ While the fidelity observation protocols that were later developed for use in the outcomes study had not yet been developed, these ratings were derived from the same criteria that the national EAST staff used to judge fidelity during the implementation study and as part of their ongoing role in monitoring program implementation.

school facilitator interviews included a total of one facilitator from each of eight and ten schools, respectively, while the student interviews included one student from each of nine and seven middle and high schools, respectively. During the second and third year of the implementation study, interviews were conducted with facilitators and students from the outcomes study target schools. In March 2005, one facilitator focus group was conducted with nine facilitators (including one backup facilitator⁷) representing all eight target schools, while two student focus groups were conducted with six students in grades 6-8 (representing three target schools), and seven students in grades 9-12 (representing four target schools), respectively. In March 2006 the facilitator focus group included the facilitator from each of six target schools (including three with middle school programs and three high school programs); and two student focus groups, one with eight students (in grades 7-9) representing all four target schools with middle school programs.

Measures/Instruments

<u>Facilitator and Student Focus Group Interviews</u>. These focus groups, which were conducted during the winter Partnership Conferences (held in Little Rock in Year 1 and in Hot Springs, Arkansas, in Years 2 and 3), were designed to help the evaluator obtain a description of the nature of the implementation of the EAST program at the school level, the contextual factors that may influence the success of the program, and the ways in which these factors vary from school to school. Interview protocols were developed to guide focus group participants through an exploration of the major research questions that the evaluation set out to study.

During the 2004 conference, four separate focus group interviews were conducted, two with facilitators and two with students, who were selected to participate as described in the preceding section. Both facilitators and students were asked to discuss the instructional methods used in the EAST classes and the strengths and challenges that they felt the program encountered, in order to determine whether and the extent to which their perspectives differed. In addition, topics that were unique to the facilitator interview included participantsø perceptions of the professional development program and the value of administrative support provided at the school, district, or state level.⁸ Topics unique to the student interview included how they became involved in the class, the characteristics of EAST students, and their opinions of their facilitatorsø teaching methods.

For the Year 2 focus groups (held March 2005) changes were made to the protocols to help deepen the evaluatorøs understanding of issues that came to light through other data collection activities. New topics added to the Year 2 facilitator focus groups included recruitment and selection of students for participation in EAST, and benefits and challenges of various grade configurations and mixes of students. Similarly, topics added to the student focus groups included studentsø technology training, the advantages and challenges of collaborating in homogenous versus mixed teams, and ways that EAST helped them to meet personal goals.

 $[\]frac{7}{2}$ Most schools employ a backup facilitator who provides support and fills in for the lead facilitator when needed.

⁸ The facilitator interview protocol also included questions addressing the nature of parental support and partnerships maintained with industry and community organizations, but these questions were omitted from the interviews due to time constraints.

Further modifications were made to the focus groups during Year 3 (held in March 2006) to spotlight additional areas that arose from other evaluation activities. The facilitator focus groups addressed the appropriateness for the EAST program of different kinds of students and further probed the previous yearøs conversation on student recruitment and selection. The student focus group also addressed new topics including their facilitatorsø teaching methods and EAST studentsø perceptions of themselves *before* they started the program in comparison to non-EAST students. This topic was included in order to gain insight as to whether there are prior characteristics needed for a candidate to succeed in EAST. Finally, students were asked to reflect on what motivates them to do well in school and on whether they felt that EAST was influencing their motivation. Copies of the interview protocols from each year are included in Appendix II.

Participantsø comments were recorded both in writing and on tape, after obtaining unanimous agreement from participants to permit the interview to be recorded (in all interviews of students and facilitators over the three years, no participant objected to recording the session). Comments were later reviewed and analyzed within each research question to describe the range of program characteristics in the participating schools and to identify the degree of similarities and differences among the schools in general, between middle and high schools and, where applicable, between facilitatorsøand studentsøperspectives.

<u>Principal Interviews</u>. Interviews were conducted during the March 2004 site visits with the principal of each of the six implementation study schools, including one interim acting principal; in two of the interviews the District Superintendent participated in at least part of the discussion along with the principal. The interview protocols were designed to obtain school-specific information about the implementation of EAST at that site and about the principalø perspectives on the program. Topics covered included the process through which the school became involved in EAST, the recruitment and selection of students for EAST classes, fiscal and non-monetary project support, professional development for facilitators, integration of EAST into the school curriculum, and parental involvement, as well as the strengths and challenges that the interviewees felt the program has encountered.

Principalsø comments were recorded in writing by the interviewer(s). As with the focus group interviews, comments were later reviewed and analyzed within each research question to describe the range of program characteristics in the participating schools, to identify the degree of similarities and differences among the different schools and between middle and high schools, and to identify any possible implementation issues that appeared to be associated with characteristics such as school size, locale, or poverty level. A copy of the interview protocol is included in Appendix III.

<u>Observations of EAST Classes</u>. During the March 2004 site visits to the six implementation study schools, school visits were scheduled to coincide with the scheduling of EAST classes, and at least two full periods of EAST classes were observed at each school. Impressions of the classes were obtained both through passive observation as well as conversations with the facilitator and the students. Observers used a narrative-based observation protocol (see Appendix IV) that focused on the physical characteristics of the classroom,

facilitator behavior and strategies, including how they organize class activities, the classroom organizational scheme and grouping patterns, the major focus of activities and materials used, and interactions with students, and studentsøbehavior and performance. These included whether the students demonstrated the ability to work both independently and collaboratively; evidence of planning and problem solving skills and strategies; and whether they demonstrated knowledge and competence about classroom procedures, project goals, and the appropriate and effective use of technology. This protocol served as both the focus of observations and a general guide for discussions with facilitators and students.

As with the principal interviews, classroom observations were analyzed to identify the degree of similarities and differences among the different schools and to identify any possible implementation issues that appeared to be associated with school level, size, locale, or poverty level.

<u>Principal Survey</u>. On-line surveys were administered to principals of all existing EAST schools in the spring of each year of the implementation study.⁹ Survey items were developed to capture issues that emerged from previously conducted school site visits, classroom observations, focus groups, and interviews, and included items addressing program funding, student and facilitator characteristics and recruitment and selection procedures, integration of EAST with other academic programs, and availability and value of administrative support. The survey was posted on the web in May 2004 and remained available for six weeks. Response rates were checked approximately once a week, and reminders were sent by ADE and/or EAST staff to those who had not responded.

In Years 2 and 3, both procedural and structural modifications were made to the principal survey, some of which were designed to help improve response rates, which were low in the first year (as discussed in the Results section below), while others were implemented to address questions that arose from the first year data and/or to provide data that specifically supported the outcomes study, which began in Year 2. Modifications designed to improve response rates included the following:

- A log-in system was created that allowed respondents to save their responses and resume where they left off so that they did not have to complete the entire survey in one sitting.
- Surveys were placed on-line earlier in the spring in order to allow more time for response.
- Sections related to the EAST training opportunities and to internal support and implementation of the EAST program at individual schools were consolidated in order to reduce the length of the survey, by prioritizing questions on a value-added basis.

In an additional effort to maximize response rates, reminder calls and follow-up emails were made to non-respondents by ADE and EAST, Inc. staff. As a result, nearly perfect response rates were achieved, as discussed in the Results section below.

⁹ In Years 2 and 3, the same surveys were also administered to the eight implementing target schools and were analyzed separately as part of the outcomes study described below.

The Year 2 principal survey was further revised to include more detailed questions related to screening criteria for EAST enrollment and new questions focused on the amount of EAST orientation and professional development activities attended by various school staff.

Finally, additional modifications were made to both the principal surveys in Year 3. Since questions about the importance of various student characteristics and their use as screening criteria were included in a supplemental recruitment survey that was administered at the outcomes study target schools (described below in section A.2.e. of the outcomes study), these items were removed from the statewide principal surveys. At the same time, since questions about recruitment and screening procedures had elicited conflicting responses from principals and facilitators in earlier surveys, question were added to both Year 3 surveys in order to clarify who has the ultimate responsibility for such decisions.

Principal surveys were analyzed to determine whether the trends that emerged during observations and interviews appeared to be representative of the state as a whole. The surveys were also analyzed to determine whether trends that appeared in the Year 1 surveys continued in Years 2 and 3. Copies of the principal surveys from each year are shown in Appendix V.

<u>Facilitator Survey</u>. A two-part on-line survey was administered to facilitators of all existing EAST schools in spring 2004. As on the principal questionnaire, survey items for the Year 1 facilitator survey were developed to capture specific issues that emerged from the observations and interviews conducted during February and March 2004. The first section included questions about the structure and implementation of the EAST program in the facilitators school, and about the facilitators opinions about factors that may be inhibiting or facilitating program implementation. This section included items that were parallel to those on the principal survey on topics covering student and facilitator characteristics, recruitment and selection procedures, and availability and value of administrative support.

The second section of the facilitator survey was originally designed to serve as a pilot survey for determining the extent to which the facilitatorsø perspectives, class organization, and instructional approach corresponded with the parameters and philosophy promulgated by the EAST national staff, and the extent to which facilitators felt prepared to implement the instructional methods expected in an EAST program. Because members of the federal review panel felt that a self-report from facilitators would not be sufficiently rigorous and would be less effective than passive observation for assessing the dynamic of the interactions between facilitators and students, they suggested that fidelity be measured instead through classroom observations at the treatment schools (as described below in the Methodology section of the outcomes study). For this reason, the section designed to assess program fidelity was removed after Year 1.

Additional modifications included questions that were added to the facilitator survey in Year 3 eliciting self-ratings of comfort using basic technologies and using the advanced technological applications that are provided through the EAST program. These items were designed to obtain insights about the extent to which some facilitatorsø concerns about the inadequacy of their technology background may have resulted from an expectationô contrary to the EAST modelô that they needed to be able to teach their students how to use the advanced technologies.

The facilitator surveys were also posted on the web in the spring of each year and remained available for six weeks, and response rates were checked approximately once a week, with reminders sent by ADE and/or EAST staff to those who had not responded. Nevertheless, as with the principal surveys and as discussed in the Results section, the response rate on the facilitator survey in Year 1 was low. In Years 2 and 3, procedural and structural modifications were made to the facilitator survey parallel to those made to the principal survey. These included the same modifications to improve response rates, as well as the substantive changes discussed above and additional substantive changes parallel to those in the principal survey. As with the principal surveys, these modifications resulted in near-perfect response rates in Years 2 and 3.

Analyses of the facilitator survey were used to determine whether the trends that emerged during observations and interviews are representative of the state as a whole, and whether these trends differ for different sub-groups (such as grade level, locale, and length of experience) of EAST schools. In addition, correlational analyses were used to explore relationships between a schooløs implementation of the EAST program and factors such as the facilitatorøs perception of the adequacy of external and internal support and funding and of the value of the program to students. Correlations with results from the principal survey were also used to determine the extent to which principals and facilitators have similar perceptions about the EAST program.

In Year 3, parallel questions to the principal survey were also added about the locus of responsibility for student recruitment and screening, as were respondentsø perceptions of their technology skills and of the adequacy of those skills for their role. Copies of the facilitator surveys from each year are shown in Appendix VI.

<u>Validity and Reliability of Locally Developed Instruments</u>. All locally developed instruments, including the class observation protocol, the facilitator and principal questionnaires, the supplemental recruitment survey, and the student and facilitator focus group and principal interview protocols, were designed cooperatively by EAST and ADE staff and the evaluator to ensure their content validity. EAST and ADE staff reviewed and edited the instruments prior to their administration to ensure their alignment with program objectives and their appropriateness for the target populations. Instruments that were administered in each project year were modified and refined, as necessary, when indicated by respondent feedback or results from prior administrations.

To ensure that respondents were familiar with the purpose of and methods for completing the instruments prior to their administration, the program evaluator provided staff with guidelines for administration of all locally developed instruments that the evaluator did not administer itself. All staff who administered the instruments were asked to adhere to standardized procedures (e.g., time frame of administration, directions given to respondents) to maximize the reliability and validity of the obtained data for evaluating the project. These procedures are described in the discussions of the individual instruments. The scoring and analysis of all locally developed instruments were conducted by the project evaluator, ensuring quality control in data entry and transcription.

Procedures

The table below summarizes which instruments and data collection methods were used to gather information relating to each research question.

Overview of EAST Statewide Implementation Study										
Research Questions	Data Sources (Years Implemented)									
	On-line Facilitator Survey (Y1-3)	On-line Principal Survey (Y1-3)	Site visits (Y3)	Student focus groups (Y1-3)	Facilitator focus groups (Y1-3)	Principal Interviews (Y1)	Observations of EAST Classes (Y1-3)	Interviews with EAST, ADE staff (Y1-3)	Observations of EAST training activities (Y1-3)	School documentation (Y1-3)
Student recruitment and screening strategies	\checkmark	\checkmark				\checkmark		\checkmark		\checkmark
Characteristics of schools, students, and	\checkmark	\checkmark	\checkmark			\checkmark	\checkmark	\checkmark		\checkmark
Training activities	\checkmark			\checkmark	\checkmark			\checkmark	\checkmark	
Instructional practices	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark			
Existing implementation measures and outcome assessments						~		\checkmark		\checkmark
Community/industry partnerships	\checkmark	\checkmark			\checkmark	\checkmark	\checkmark	\checkmark		
Parent involvement programs		\checkmark	\checkmark			\checkmark	\checkmark	\checkmark		
Administrative support from school, EAST, and ADE staff	~	\checkmark			\checkmark	\checkmark		\checkmark		
Dissemination activities and capacity building initiatives		\checkmark				\checkmark		\checkmark		
Program costs		\checkmark						\checkmark		\checkmark
Obstacles to implementation	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		

Table 2	
Overview of EAST Statewide Implementation	Study

Results

Of the 134 EAST programs in schools of the same structure as the schools that would be targeted for the outcomes study (public middle and high schools) that were operating during the 2003-2004 school year, responses were received from a total of 62 principals, for a response rate

of 46%.¹⁰ Subsequently, surveys were received from the principals of all 133 public middle or high school EAST programs in SY 2004-2005,¹¹ and from all of the applicable 134 EAST programs operating in SY 2005-2006, for response rates of 100% in each of the last two years.

Of the 134 EAST programs operating during 2003-2004, responses were received from a total of 77 facilitators, for a response rate of 54%.¹² In Years 2 and 3, respectively, surveys were received from the facilitators at 121 out of 131 EAST programs¹³ for a response rate of 92% in SY 2004-2005, and from 119 out of 134 EAST programs for a response rate of 89% in SY 2005-2006. Results of the facilitator and principal surveys are summarized below. Additional data from these surveys can also be found in Appendices VII-X.

Student Recruitment and Selection Strategies

While the EAST program is not intended to be appropriate for every student who comes through the school, it is considered to be beneficial for a very diverse cross-section of students; indeed, diversity is considered an important characteristic of an EAST class and is considered as a goal towards which programs should strive. The approaches which different schools take for recruiting and selecting students for EAST classes is therefore an important program characteristic, both in terms of the extent to which they are following the model and of the implications of these practices for making generalizations about the value of the program. Since in virtually all schools, the principal and the EAST facilitator shared responsibility for recruitment and selection of students, questions about these policies were included on both surveys.

When asked what methods were used to recruit students for EAST, both principals and facilitators in all three study years reported a combination of strategies that promoted general awareness of the program and those that targeted specific students. For example, substantial majorities of responding principals in all three years indicated using all-school announcements to recruit students. While this approach was also cited by large proportions of facilitators, it was cited less frequently by facilitators than by principals, and the proportion doing so declined steadily (from three fifths of respondents to just under half) from 2004 to 2006.

Strategies such as this which cast a wide net are important for fostering diversity in the EAST classes, since they help ensure that all students are aware of the program. However, they

¹⁰ Principal surveys were only administered to the population of schools in Arkansas that had an active EAST program during that year and that were comparable in structure to the study schools. This excluded colleges, technical colleges, community colleges, vocational centers, and elementary schools that were also conducting EAST programs. In spring 2004, one additional survey was received from the Southeast Arkansas Community Based Education Center (SEACBEC), which should not have been included in the survey.

¹¹ An additional 12 EAST programs sponsored by other types of agencies also returned principal surveys in spring 2005. These included North Arkansas College, Fountain Hill Elementary School, Aerospace Education Center, and Acxiom Corporation, and six surveys from schools with EAST programs that had been discontinued, including Southside Middle School, and Caddo Hills, Har-ber, Jacksonville, Saratoga and Whitehall high schools.

¹² As for the principal surveys, additional surveys were received from other agencies that should not have been included in the survey. These included North Arkansas College, SEACBEC, and the Aerospace Education Center ¹³ An additional five EAST programs sponsored by other types of agencies also returned facilitator surveys in spring 2005. These included North Central Vocational Center, North Arkansas College, SEACBEC, Fountain Hill Elementary School, and Acxiom.
do not necessarily result in classes that are *representative* of the school, since they tend to rely on the students to take the initiative to respond. Also in 2005 and 2006, the large majority of facilitators (78% to 90%) and over three quarters of principals reported the use of õpeer recruitment,ö where the program is advertised among students by word-of-mouth.¹⁴ The impact of this strategy on diversity is unpredictable, but it might be expected to reinforce existing enrollment patterns, whether existing enrollment favors certain groups or is already well diversified.

The self-selection that tends to result from such broader recruitment practices can be counterbalanced by more targeted strategies, but only if the recruitment is specifically targeted towards the students who would be less likely to enroll on their own. In all three years, approximately 7 to 9 out of every 10 principals, and similar majorities of facilitators, noted that school staff (including guidance counselors, facilitators, classroom teachers and/or principals) recommend students for EAST. However, the impact of these targeted strategies depends on the criteria that are used to determine which students should be recommended for the program, which can mean the difference between counterbalancing or exaggerating the self-selection bias. This issue was explored further through the supplemental recruitment and screening survey that was administered during the outcomes study, and is discussed further in that section.

Although the survey asked about school-wide recruitment practices, responses from principals and facilitators were not always aligned with each other. For this reason, further analyses were conducted to examine the level of agreement between their responses. The only recruitment strategies which were cited by both the principal and the facilitator at more than half the schools in both years were peer recruitment, recommendations from the facilitator, and recommendations from the principal or guidance counselor. Recommendations from teachers were also cited by both the principal and facilitator at just over half the schools on the 2005 survey.

Both principals and facilitators were also asked how students were selected to participate in EAST. In all three study years, both principalsø and facilitatorsø most common response (cited by about three fifths of both principals and facilitators) was that preference was given to students who have participated in EAST in previous years. The second most commonly cited approach was that students are chosen in order to reflect the range of academic abilities in the school, which was cited by about half of the principals. Almost as many facilitators (about two fifths of respondents) also mentioned this approach in 2004, but very few (about one-fifth) did so in 2005 or 2006. Over the three years, among the most commonly cited responses from both groups was õfirst come first served,ö cited with increasing frequency in each year by two fifths to over half of principals, and by one third to over two fifths of facilitators. Substantial proportions of principalsô although not a majority (ranging from about one third to two fifths of respondents) also cited efforts to reflect the ethnic makeup of the school, but only about one fifth to one third of facilitators did so.

As for the recruitment questions, separate analyses of agreement between facilitators and principals were conducted in 2005 and 2006 in order to shed light on apparent discrepancies between these two respondent groups. In both years, the strategies which facilitators and

¹⁴ This option was not included on the 2004 surveys.

principals agreed to using in the largest proportions of schools (albeit in only just over one school out of four) were preference given to students who had participated in previous years, and first come, first served. While the former strategy is consistent with the model (since õseedingö the class with experienced EAST students is believed to help the program run more smoothly), the first come, first served approach would do nothing to insure diversity or representativeness. In contrast, there was agreement between the principal and facilitator in only very few schools about using efforts explicitly designed to reflect the school¢s range of academic abilities and ethnic makeup, while the two groups agreed that such efforts were *not* used in about half or just under half of the schools in both years. The generally low degree of alignment between principalsø and facilitatorsø responsesô which on many items were consistent at fewer than 60% of the schoolsô raises questions about what actually took place.

In an effort to clarify that question, the 2006 surveys asked who was responsible for recruitment and selection, and who had authority for making the final decision. While almost all principals (over 90%) asserted responsibility for advertising EAST, most facilitators did not seem to recognize this, and indeed there was agreement that the principal had such responsibility in fewer than one out of every six schools. Similarly, a smaller majority of principals (75%) reported having responsibility for selection and placement, but there was agreement with the facilitator about this in only one in every five schools. In addition, in quite a few cases the facilitators believed that they possessed responsibility but were contradicted by their principals. This occurred at over half of the schools in reference to facilitatorsø responsibility for recruitment, and at almost a fourth in reference to responsibility for selection. There was a much higher degree of agreement about the school counselor(s) having responsibility for recruitment and screening (principalsø and facilitatorsø responses were in alignment at two thirds to three fourths of responding schools), and other than principals, counselor was the role most commonly cited by either group as having this responsibility. However, when asked who makes the final decision about student placement, each group chose the other most frequently, with principals emphasizing the facilitators@responsibility (68.6%), and facilitators emphasizing the principals@ responsibility (41.9%).

Detailed results of survey questions regarding recruitment and selection are presented in the tables in Appendix VII.

Training and Orientation Activities

EAST facilitators participate in ongoing training activities throughout their tenure as a facilitator, beginning with their first introduction to the initiative during Vision Building workshops. The required professional development activities are summarized in the table below.

Session	Schedule	Focus				
Vision Building	Spring before becoming	Orientation to basic concepts, goals, and philosophies of				
	an EAST site	the EAST program				
Phase I	Summer before first	Logistical issues, team building, EAST vision and				
	year as facilitator	philosophy, learning styles, roles of students and				
		facilitators, grading and assessing, project planning,				
		technology and resources, pitfalls				
Phase II	October of each year	Sharing, traps and problem identification, creating service				
		projects, professional development, technical support				
Phase III	December of each year	Reflection on Phase II, corporate culture, marketing				
		EAST, equity, growing the program				
Summer Seminar	Summer of each year	Reflections on the year, using resources, planning for next				
		year, new tools for the classroom, sample projects				
EAST Updates	Fall of each year	Training policies and procedures, conference updates, use				
		of the website, planning õsite healthö visits, network				
		security, computer maintenance, obtaining technical				
		support				
Partnership	February of each year	Student presentations and awards, vision building, web				
Conference		support				

Table 3Professional Development Activities for EAST Facilitators

In addition to the above, program students have access to ongoing training related to the technologies available in EAST classrooms, which is provided by representatives of the software companies and EAST staff, and through turnkey training from lead students in their class. Students may also participate in peer visits, during which they visit another EAST class in order to learn from and network with other EAST students (as well as to collaborate on cross-class or cross-school projects).

Principals, other school administrative staff, and other members of the school community also receive orientation to the EAST program through New Administratorsø Workshops, which are generally held in the fall of each year, and Pre-Implementation Workshops, which are held in July before a school¢ first year in EAST, in addition to the Vision Building Workshops and Partnership Conferences described above.

Each year, EAST facilitators and principals were asked to report on their perceptions of the impact of these training activities. Results from the 2004 and 2005 surveys were presented in previous reports. On the 2006 survey, principals and facilitators continued to express very positive feelings about the training process. On a 5-point scale ranging from 1 (*inhibits effective implementation*) to 5 (*encourages effective implementation*), about half of principals and similar proportions of facilitators rated õstaff training sponsored by EAST, Inc.ö and õstudent training sponsored by EAST, Inc.ö as a 5, and about three quarters of each group rated both trainings as at least a 4. Similar proportions of principals and facilitators also rated õEAST conferences and competitionsö as encouraging effective implementation.

Facilitators were also asked to rate the adequacy of EAST training for bringing their skills to the level they needed as a facilitator. On a 5-point scale ranging from 1 (*inadequate*) to 5 (*more than adequate*), facilitators reported that the training was best for raising their skills in the instructional methods advocated by EAST, which was rated as at least *adequate* by

approximately four fifths of respondents. Facilitators also provided favorable ratings to training they received in assessing their own progress in facilitating EAST, which was rated as at least *adequate* by three fourths of respondents. The lowest ratings were for training in the specific technology applications used in the lab, for which fewer than half of respondents felt the training was at least *adequate*. Facilitators were also asked to rate their own skills in using both basic and EAST-specific technologies as *novice*, *beginner*, *intermediate* or *advanced*. While almost all respondents (over 90%) rated their basic technology skills as at least *intermediate*, almost two-fifths felt that their comfort and skills with specific technology applications used in the EAST classroom were at the beginner or novice level. As in past years, facilitatorsø overall ratings of EAST trainings were fairly positive and consistent. On a 7-point scale (1 = *poor* and 7 = *excellent*), the greatest portion of respondents rated the training as a 6, with approximately 9 out of 10 ratings falling at 5 or above.

Detailed results of survey questions regarding training activities are presented in Appendix VIII.

Program Support and Obstacles to Implementation

School-based staff have made clear that they consider administrative supportô from the school and district level as well as from the EAST National Center and the ADEô as essential to the success of EAST. In order to assess the adequacy of support available to EAST programs, both principals and facilitators were surveyed about their impressions of the support that they received, while principals were also surveyed about their opinions about establishing supportive school policies and practices. Results of the 2004 and 2005 surveys have been presented in a previous report.

As in past years, both facilitators and principals were asked to rate the impacts of various external program supports on EAST implementation in their schools. On a 5-point scale ranging from 1 (*inhibits effective implementation*) to 5 (*encourages effective implementation*), facilitatorsøratings of supports other than training (discussed above) were mixed. Availability of technology support was the highest rated, with approximately three quarters of respondents rating it as a 4 or more and almost a third rating it as a 5. However, the only other forms of support that were rated as at least a 4 by at least half of respondents were availability of funding to support participation in EAST training, opportunities for the facilitator to work with other EAST facilitators, availability of support on logistics of program implementation, consistency of EAST philosophy with other school/district reforms, and time for facilitator(s) to reflect on what they have learned. In contrast, state-mandated curriculum frameworks, testing policies and practices, and grading policies and practices were viewed as supporting implementation by one fifth or fewer of the respondents, but were seen as *inhibiting* implementation by about the same proportions. Testing policies were seen as the greatest obstacle, which 3 in 10 respondents rated as a 1 or 2. Ratings from principals demonstrated similar patterns.

Facilitators were also asked about the adequacy of õinternalö supports that they received from their school administrations. Over 90% of respondents agreed that their principals supported creation of classes with students of mixed abilities and that they were able to obtain permission for EAST students to work off campus when necessary, while 85% agreed that their principals encouraged other teachers to be flexible so that EAST students could make up missed

classes. A large majority (about three-quarters) also agreed that their principals facilitated class scheduling in order to involve as many interested students as possible in EAST, but a smaller majority (a little over three-fifths) reported that their principals facilitated class scheduling in order to place experienced EAST students in the labs.

The principal surveys asked these school leaders about their willingness to provide certain kinds of direct support to the EAST classes, as well as their perceptions about various instructional methods and philosophies that are central to EAST. Principals echoed the facilitatorsø impressions about their willingness to give EAST students flexibility to make up missed classes and to go off campus, and to facilitate class scheduling. In addition, there was also strong agreement about almost all of the other EAST philosophies, with the strongest agreement (from 90% of respondents or more) that using academic skills in a real-world context is important to student learning, encouraging students to solve their own problems is more important than making sure they find the right answer, and that most students benefit from working in groups to solve problems. The only notable exception was that a small majority of principals (just over half) felt that most students learn best in classes with students of similar abilities. This was a somewhat surprising finding, given that the large majority of facilitators had said that they were supported in creating mixed-ability classes.

In order to determine whether EAST functioned as an integral part of the school curriculum or more as an õadd-on,ö principals were asked in what ways (if any) EAST was integrated with other courses or programs at their schools. Each of the forms of integrationô including those reflecting EAST supporting students in other activities or subjects, as well as those reflecting other activities or subjects supporting students ø performance in EASTô was cited by a substantial majority of respondents. Virtually all principals agreed that *EAST students contribute to other school activities*, and a large majority (over 80%) felt that *skills learned in other subjects help students succeed in EAST*. Substantial but somewhat smaller majorities (about three fourths of respondents) felt that *papers or projects for other classes may be based on topics from EAST projects, EAST projects incorporate specific content knowledge (besides technology) that is covered in other courses, and that EAST develops students' general academic habits and attitudes. The smallest majority (two thirds) of respondents reported that EAST develops specific academic skills in the content areas that are applicable to other classes.*

Detailed results of survey questions regarding program support and obstacles to implementation are presented in Appendix IX. Additional data from the principal and facilitator surveys are presented in Appendix X.

Methods

Participants

Selection of Experimental and Control Schools

In preparation for the outcomes study, which began in Year 2, ADE and EAST staff collaborated with the evaluator in April 2004 to identify schools that would participate in the study as experimental (õimplementingö) and control (õdelayed implementationö) schools.

In order to insure sufficient statistical power for treatment/control comparisons in this design, a power analysis was conducted to determine the minimum number of students required. This analysis provided our best *a priori* estimate for the smallest sample size needed to detect a predetermined effect size. The following sample size calculations are based on conducting an Analysis of Variance (ANOVA), with mean values of each outcome measure, compared for treatment vs. control students in different contexts, including different school locales and school levels.

Experimental conditions were defined within four õcellsö (two locale categories by two levels, middle and high school). The original plan was to identify two implementing and two delayed implementation schools for each of the eight cells, for a total of 16 schools, assuming that at least that many schools volunteered to join the EAST initiative.

Because the ANOVA design reduces extraneous variation by controlling for these contextual conditions, it is capable of providing substantial statistical power with relatively modest sample sizes. The minimum number of cases per group was determined from Table 8.4.4 in Cohen (1988, pg. 384), which provides minimum sample sizes necessary to detect specific effect sizes using a non-directional F test on means in ANOVA at alpha = .05, for different levels of power and degrees of freedom. In order to detect an effect size of approximately one-third standard deviation (0.30) with power (beta) of 0.80 and one degree of freedom (u), the table indicates that a minimum of 45 cases are needed per cell. It should be noted that this analysis is predicated on two additional assumptions:

- 1. control and experimental groups have equal sample sizes and
- 2. control and experimental group means have equal standard deviations.

We insured that the first assumption was met by matching participating students in each treatment school to an approximately equal number of students in each control school (as described further below). Given similar sample sizes, it was considered unlikely that standard deviations would vary substantially, but it should be noted that power determinations are robust to moderate violations of equality of standard deviations and sample sizes (Cohen, 1988).

According to these calculations, a sample size of 45 treatment students plus 45 control students per sample group would provide sufficient statistical power to detect an effect size of 0.30. However, by attempting to improve the power to detect the effect size of treatment vs.

control groups by matching students, an n of 45 was expected to produce the ability to detect even smaller effect sizes with equal power.

Because of the significant expense of establishing an EAST class, schools rarely if ever have more than one lab room, but they do tend to make full use of their lab. It is therefore typical that even the smallest schools participating in this project serve at least three separate EAST classes per year, which was expected to yield approximately 45 to 60 participating students per school. Spread out over six grades (7 to 12) and two locale conditions, a sample size of eight experimental and eight control schools was expected to yield at least 35 students in each group.¹⁵ In order to make it possible to collapse analyses across grades in the event that the final sample size was too small to analyze the data within grade, outcome instruments were chosen that permit scores to be combined across grade levels. Ultimately, our final sample of target students included 412 EAST students in Cohort 1¹⁶ and 509 in Cohort 2.¹⁷

Prior to the first year of the outcomes study, EAST staff discussed the evaluation with schools that were considering adopting an EAST program. During the spring of each school year, EAST, Inc. conducts õVision Buildingö workshops for such schools, during which EAST staff explain the expectations of becoming an EAST school. During the spring 2004 workshops, staff also explained the implications of joining the program during the study period. As an incentive for schools to take the risk of being selected as a control school and required to delay program implementation for two years, ADE agreed to set aside start-up funds that were made available to delayed implementation schools after the end of the study period. As an additional incentive during the two years of the outcomes study, and to minimize õresentful demoralization biasö¹⁸ among control schools, the state had further agreed not to provide funding to support the establishment of an EAST program during the study period at any school that did not participate in the study.

Following the Vision Building workshops, 23 schools expressed interest in establishing an EAST program in the following year, but three schools that had already secured EAST funding in the prior year decided to opt out of the study by funding their program independently.¹⁹ EAST staff provided the evaluator with a list of the remaining 20 schools that still expressed interest in implementing the program starting in the 2004-2005 school year. These 20 schools were grouped into one of four categories defined by the grade levels served (middle or high school) and the school¢s locale (rural or urban, as defined above for the

¹⁵ Assuming that the typical new school has at least three EAST classes with 15-20 students each, i.e., at least 45-60 students. Assuming a midpoint of 52 students per school x 8 experimental schools x 6 grades x 2 locale categories = 34.7 students per grade per locale category.

¹⁶ Including 208 in rural high school programs, 104 in urban middle school programs, and 100 in rural middle school programs.

¹⁷ Including 184 in rural high school programs, 164 in urban middle school programs, and 149 in rural middle school programs.

¹⁸ After Campbell & Stanley (1966).

¹⁹ Although permitting schools to opt out of the study prior to randomization may create a selection bias, prohibiting schools from opting out might have created an equally large selection bias if some schools decided that they did not want to implement the program at all, given the studyøs requirements. EAST and ADE staff decided that it was preferable to maximize replication of the program by adhering to the former approach.

implementation study target schools²⁰). Consistent with the February 2004 Cooperative Performance Agreement with the USDOE, the study was to include two implementing and two delayed implementation schools in each of these four categories or õcells.ö However, the original list of schools applying for EAST included only one rural middle school and two urban high schools, and EAST and ADE staff did not expect to be able to recruit a sufficient number of additional new schools to fill these gaps. At ADEøs request, three rural schools serving grades 7 to 12 that had originally intended to implement the program at the high school level agreed to implement EAST in their middle school grades for the duration of the study.²¹ Since there was still an insufficient number of urban high schools on the list, however, this cell was dropped from the outcomes study.

From each of the middle school cells (rural and urban), two schools were randomly selected as implementing and two as delayed implementation schools. However, because urban high schools were dropped from the study, it was necessary to select four implementing and four control schools from among the rural high schools, in order to retain sufficient statistical power for analyses at the high school level. Fortunately, this category had the largest number of applying schools, making this larger sample possible. Since this category also represented the largest proportion of existing EAST schools, this modified sample remained consistent with the population. The two remaining (non-participating) schools in this category (the only category that had more applying schools than were needed for the study) and the two urban high schools that did not participate because of the change in the sample definition were told that they were free to implement EAST using their own resources. They elected to wait for EAST funding, however, and were also able to implement the program with the support of ADE and EAST funds after the study ended.

In spring 2004, as a result of state-mandated mergers of certain school districts throughout Arkansas, the district in which Evening Shade High School (one of the original implementing schools) resided was administratively consolidated with the Cave City School District. As a result, the fledgling EAST program that was to be implemented at Evening Shade was moved to Cave City High School, which replaced Evening Shade as an implementing school for the outcomes study. Despite the obvious concerns raised by this move, it seemed preferable to losing a school entirely so early in the study. Several considerations added to the expectation that this replacement should not unduly bias the study. These included the fact that both schools had the same NCES locale code, so the replacement preserved the balance between rural and urban schools in the study; the principal of Cave City had participated in the Vision Building workshops during summer 2004, during which he received full orientation to the program and to the obligations of the study, and he readily agreed to sign the Statement of Commitments regarding those obligations; and finally, the EAST facilitator who was originally identified by Evening Shade transferred to Cave City and implemented the EAST program there.

²⁰ Because the state of Arkansas has very few students who are served by schools located in or near large metropolitan regions, the PET had determined that the target population for the experimental study could be constrained to include only two locale categories (small city/large town vs. small town/rural, as defined under the implementation study above), rather than the three originally planned (large city/small city, large town/small town and rural).

²¹ These schools included Harrisburg, Gould, and Hartford high schools.

Similarly, Gould High School (which served grades 7 through 12 and had an EAST program for students in grades 7 through 9) was annexed to Dumas Junior High School as of fall 2005, and the EAST program continued in Dumas with the same facilitator. As was the case with the replacement of Evening Shade by Cave City, Dumas was in the same district and had the same NCES locale code as Gould, and the program was served by the same facilitator in both locations. However since the principal of Dumas was new to EAST, the ADE Director of Technology Resources and Planning met with him to orient him to the program, and the district technology specialist worked with the principal and the EAST facilitator to help get the program up and running quickly.

The resulting sample of schools for the outcomes study is displayed in Table 4 below.²²

	Demograph	Lovel of		tug	biuuy	
	Locala		Veen 2	Veen 2 Veen 2		
School Nama	Cotogory ²³	LASI Classos	Y ear 2 (SV 2004 05)	Y ear 5 (SV 2005 06)	Grades	Enrollmont*
School Name			(SY 2004-05)	(SY 2005-00)	Jerveu	
Blevins HS	Rural	HS	Implementing	Implementing	7-12	1 Quartile
Newport HS	Rural	HS	Implementing	Implementing	9-12	3 rd Quartile
Jessieville HS	Rural	HS	Implementing	Implementing	7-12	3 rd Quartile
Cave City HS		HS	Implementing	Implementing	9-12	1 st Quartile
(originally Evening	Rural					
Shade HS)						
Midland HS	Dunal	HS	Delayed	Delayed	7-12	2 nd Quartile
	Kurai		Implementation	Implementation		
Mt. Vernon-Enola	D 1	HS	Delayed	Delayed	7-12	1 st Quartile
HS	Rural		Implementation	Implementation		
Mountain Pine HS	. .	HS	Delayed	Delayed	7-12	4 th Ouartile
	Rural		Implementation	Implementation		
Mansfield HS	D 1	HS	Delayed	Delayed	7-12	4 th Quartile
	Rural		Implementation	Implementation		
Henderson MS	Urban	MS	Implementing	Implementing	6-8	4 th Quartile
North Heights JHS	Urban	MS	Implementing	Implementing	7-8	4 th Quartile
Cloverdale MS	Linhan	MS	Delayed	Delayed	6-8	4 th Quartile
	Urban		Implementation	Implementation		
Forest Heights MS	T 7 1	MS	Delayed	Delayed	6-8	4 th Quartile
	Urban		Implementation	Implementation		-
Harrisburg MS	Rural	MS	Implementing	Implementing	5-8	2 nd Quartile
Gould HS	D 1	MS	Implementing	[annexed to	7-12	1 st Quartile
	Kurai			Dumas]		-
Dumas JHS	Rural	MS	[NA]	Implementing	7-9	3 rd Quartile*
Flippin MS	Dunal	MS	Delayed	Delayed	6-8	1 st Quartile
	Kurai		Implementation	Implementation		-
Hartford HS	D 1	MS	Delayed	Delayed	7-12	1 st Quartile
	Kural		Implementation	Implementation		

 Table 4

 Demographics of Target Schools for Year 2-3 Outcomes Study

*Enrollment quartile for Dumas based on SY 2005-2006 enrollments; all other schools based on SY 2004-2005 enrollments.

²² Enrollment groupings were defined by identifying quartile distributions of all EAST schools in Arkansas for the number of students enrolled in each school during the 2002-2003 school year.

 $^{^{23}}$ õUrbanö = NCES codes 2-5; õRuralö = NCES codes 6-8.

Schools that were selected as implementing or delayed implementation schools were asked to sign a Statement of Commitments prepared by the evaluator,²⁴ in addition to the Statement of Assurances which all EAST schools are normally required by EAST, Inc. to sign. The Statement of Commitments, which all selected schools agreed to sign, committed the delayed implementation schools to postpone implementation of EAST until after the study period, regardless of funding sources, and to participate as control schools in selected evaluation activities during the study period. Implementing schools committed to conditions specifying selection of their facilitator and the number and size of their EAST classes, and to participate in all evaluation activities during the study period.

Selection of Control Students

Two slightly different selection processes were used in each year of the outcomes study to create the control group. In the first year of the study (school year 2004-2005, Cohort 1), all EAST students in the participating implementation schools were included in the treatment group for the study, while a sample of students was selected from within the delayed implementation schools in order to create a control group that was as closely matched to the treatment group as possible. This sample was selected through pairing on the basis of prior year academic performance and demographic factors.²⁵ The specific pairing variables that were used, and the source of data for each variable, are presented in the table below. Matching criteria are listed here in descending priority order, with higher priority variables taking precedence, when necessary, in the pairing algorithm (explained further following this table).

²⁴ A copy of the Statement of Commitments is provided in Appendix XI.

²⁵ Demographic data were obtained from ADE databases for the 2003-2004 school year (as of June 2004), since fall 2004 records were not yet available when control students were being selected.

Control Group Matching Criteria Conort I					
Attribute	Data Source ¹				
School locale	NCES records (õruralö/õurbanö ²⁶)				
Program level	EAST, Inc. records (middle/high school)				
Grade level	ADE records				
English proficiency	ADE records (ELL flag)				
Gender	ADE records				
Reading proficiency	Spring 2004 standardized test scores:				
	ITBS/ITED Reading (grades 6 and 9);				
	Arkansas Benchmarks Exam (grade 8);				
	Arkansas 11 th Grade Literacy Test (grade 11)				
Math proficiency	Spring 2004 standardized test scores:				
	ITBS/ITED Math (grades 6 and 9);				
	Arkansas Benchmarks Exam (grade 8)				
General academic achievement ²⁷	ADE records (gifted/talented and special education flags,				
	grades 5, 7 and 10)				
Race	ADE records				
Socioeconomic status	ADE records (free/reduced lunch eligibility)				

Table 5Control Group Matching Criteria – Cohort 1

After collating the above data for all students enrolled in the delayed implementation (control) schools in fall 2004, the students were ranked around the criteria for each EAST student and selected for proximity. Control groups were selected that contained the same number of students in each school type and grade as in the treatment groups. In addition, alternate control students were selected to be used as replacements in the event that any of the students in the primary control group were no longer at that school. Within each grade and school type, alternate control students totaling at least 10% of the N for that groupô with a minimum of two studentsô were selected.²⁸ Alternate control students were selected on identical criteria, but were selected to be as close as possible to the mean of the standardized test score(s) for the corresponding treatment group, since the mean would represent the best possible estimation of the value for a student who was being replaced. Where possible, a mix of demographics (through alternating selection) was also represented within a group of alternate control students. Further provisions had also been made for matching by distribution and by distance metrics derived from these attributes; however, these procedures were unnecessary, as excellent matches were obtained using simple averaging of distance of standardized test scores within combined demographic categories. Results of these matching procedures are presented in Appendix XII, which shows the demographic characteristics of the treatment and resulting control groups. These tables represent all control and alternate control students who were selected through the pairing process. Analyses of the impact on the makeup of the target groups from demographic differences between students who left the study (school transfers or dropouts) and those who remained are discussed below.

²⁶ As defined above for the implementation study target schools.

²⁷ Since Arkansas did not conduct math and literacy testing at all grades in spring 2004, gifted/talented and special education flags were used as surrogate measures of academic achievement for students at grade levels that were not tested, but were also included as matching criteria for the remaining students.

²⁸ At one school, as a result of inaccuracies in student enrollment records, a substantial number of the control students from the primary selection group were no longer enrolled at the school, and there was not a sufficient number of alternate control students from the original selection to fill the gaps. In this case, the selection procedure for alternate control students was repeated in order to provide a sufficient number of students.

The control group for Cohort 2 (students attending Arkansas schools during school year 2004-2006) was created using essentially the same procedure. The primary differences were that the delayed implementation schools had decided that it was easier to administer achievement tests, inventories, and surveys²⁹ to the whole school rather than pulling out a selected population for testing. Thus, the Cohort 1 control group had been created at the beginning of the school year based on prior year performance and included extra students who could be used in case some students were not available or in case of missing data. In contrast, the Cohort 2 control group was matched after the fact using performance and demographic factors from the fall of the same vear. Wherever possible, this matching used only those students who had a pretest available on the key outcome measures, which included the Iowa Tests of Basic Skills (ITBS) and Iowa Tests of Educational Development (ITED) (used as the state achievement tests) for middle and high school students, respectively, and inventories of motivation and problem solving skills. In addition, whereas the spring 2004 data for Cohort 1 included ITBS/ITED scores only for grades 6 and 9, by fall 2005 ADE had begun testing all students in grades 6 though 9, and, as part of the study, also administered the ITED in grades 10-12 among EAST students in the implementing schools and all students in the delayed implementation schools. Although matching was conducted after the end of the school year, spring posttest scores were not visible to the individuals conducting the matching, so as to avoid the possibility of creating a bias in the selection process. Matching criteria for Cohort 2 are listed in the table below in descending priority order, with higher priority variables again taking precedence in the pairing process.

Attribute	Data Source					
School locale	NCES records (õruralö/õurbanö ³⁰)					
Program level	EAST, Inc. records (middle/high school)					
Grade level	ADE records					
English proficiency	ADE records (ELL Y/N)					
Gender	ADE records					
Reading proficiency	Fall 2005 ITBS/ITED Reading Comprehension					
Math proficiency	Fall 2005 ITBS Math Total ³¹ /ITED Concepts & Problems					
General academic achievement	ADE records (gifted/talented Y/N and special education Y/N)					
Race	ADE records (Asian, Black, Hispanic, Native American,					
	White/non-Hispanic)					
Socioeconomic status	ADE records (free/reduced lunch eligibility Y/N)					

Table 6 **Control Group Matching Criteria – Cohort 2**

Statistical comparisons of the resulting demographic characteristics of the Cohort 2 treatment and control groups were conducted in order to determine the success with which the matching process created a comparison group that was demographically similar to the treatment group. Demographics of the two target groups are presented in Appendix XIII, along with results of the statistical comparisons. These comparisons confirmed that the resulting control group was statistically indistinguishable from the target EAST group in terms of their ITBS reading and math pretest scores, gender, grade level distributions, or proportions of gifted and

²⁹ See summary of instruments in the next section.
³⁰ As defined above for the implementation study target schools.

³¹ Excluding the Computation section.

special education students and English language learners (ELLs).³² However, there were significant differences between the two groups in terms of racial distributions and proportions of Title I eligible students, with the control group having slightly larger proportions of non-minority students (73% vs. 62%) and smaller proportions of African-American students (22% vs. 33%),³³ and slightly larger proportions of Title I eligible students (90% vs. 84%).³⁴ Although these differences were statistically significant, the magnitudes of the differences between the groups were modest and seem unlikely to have a meaningful impact on the outcomes.

Measures/Instruments

Logic Model

A detailed definition of the EAST program can be found in the logic model presented in Appendix I. In this table, the first column (EAST Program Components) describes in detail the various characteristics of the EAST model, within each of the three environments in which EAST is conceived to take place: the physical environment, the educational environment, and the environment of expectations. Included among these program components are descriptions of the ideal classroom setup, expected instructional methods and class and program management strategies, characteristics of student projects and work conditions, and classroom atmosphere or culture. Although instructional methods and instructor strategies can be thought of as outcomes that are expected of the facilitators (as a result of training), they are included under program components because they are also among the aspects of the EAST model that are believed to make a critical contribution to student outcomes. The second column of this table (Student Outcomes) describes the outcomes that the EAST program is believed to promote in participating students. These are described in terms of the knowledge, skills, intellectual abilities, and attitudes that the program is believed to develop. After each bullet in the Program Components section is a number or list of numbers that refer to the specific outcomes (from column 2) to which that component is believed to contribute.

Column 3 (Outcome Variables) of this table shows a selected list of the variables that were examined during this study, horizontally aligned with the outcomes to which they relate. Each variable relates to the outcome with the same number that appears in the same row in column 2. Finally, the last column (Measurement Instruments) indicates the measurement instrument that was used to assess each outcome variable. Each measure relates to the variables and outcomes in the same row in columns 2 and 3.

While the EAST program theory had already been well established at the beginning of the study, it had never been explicitly defined, nor had the linkages between program

³² ITBS Reading: Independent Samples *t*-test t = .416 p = .677

ITBS Math: Independent Samples *t*-test t = .492 p = .623

Gender: Pearson Chi-square =.093, df = 1, p = .761

Grade level: Pearson Chi-square = 2.378, df = 6, p = .882

Gifted: Pearson Chi-square = .048, df = 1, p = .847

Special Education: Pearson Chi-square =1.958, df = 1, p = .162

ELL: Pearson Chi-square = .017, df = 1, p = .898

³³ Pearson Chi-square = 15.502, df = 4, p = .004

³⁴ Pearson Chi-square = 11.198, df = 1, p = .001

components and outcomes. In order to better document these concepts and to help the evaluator develop a clearer understanding of the EAST model, a logic model was developed through collaboration between Metis, EAST, and ADE staff. Creating the model was a complex and time-consuming process that involved extensive conversations between key stakeholders that necessitated repeated, iterative cycles of revisions in order to insure that it accurately reflected program theory and operationalized that theory in objectively measurable terms.

The logic model was used throughout the evaluation to facilitate discussions on program capacity and growth as well as to develop instruments used in defining fidelity to the program (observation protocol), student satisfaction and perceptions (student survey), and recruitment and selection processes by school (recruitment survey). Perhaps most importantly, development of the observation protocol relied heavily on the logic model in order to help create an indicator of program fidelity by school and facilitator. Program components as defined in the logic model served as the first step in identifying the types of behaviors and activities which the observation protocol, including the annotated version, which provided more specific information, clarification, and examples as to what constituted high fidelity and what was expected from an effective EAST program. This information was invaluable to observers in the assigning of ratings (both overall and individual area) during observations.

By making the program theory explicit, the logic model has proven valuable not only to the evaluation, but has become a living document which EAST staff continue to use to support their professional development and dissemination initiatives. EAST staff are using the model to help familiarize school-based staff with program aspirations and goals, and to support their own õsite healthö observations through which EAST, Inc. staff monitor and support program implementation. The process overall was universally seen as useful, effective, and crucial to the further development of the program and evaluation.

Classroom Observations

The model of an EAST program incorporates a broad range of dimensions, from the physical layout of the classroom, to the strength of the schooløs connections to the community, to the facilitatorøs expectations of each studentøs ability to succeed.³⁵ But it is the facilitators themselves and how they run their EAST classroomsô their approach to instruction, the nature of their interactions with studentsô that represent the most crucial aspects of the model and are the most important for obtaining fidelity of program implementation. For this reason, observations of EAST classes were conducted as one of the primary measures of the fidelity with which the EAST program was being implemented in each school, with a focus on whether each facilitator was following what is expected of an õideal facilitator.ö

In order to standardize the assessment of classroom implementation, the program components described in the logic model table were used as the basis for developing a classroom observation protocol that was also supported by an Annotated Guide designed to help clarify and operationalize the concepts that observers were asked to rate, and by Fidelity Observations Guidelines, which detailed the procedures which observers were expected to follow during the

³⁵ These dimensions are listed in the Program Components column of the logic model in Appendix I.

training, pilot testing, and formal observation phases. (All three documents are shown in Appendix XIV.) Trained observers used this protocol to rate the extent to which the observed classes and classrooms demonstrated each of 84 specific characteristics indicated by the logic model. Each characteristic was rated on a 5-point scale, where 1 indicated that the characteristic was õnot at allö demonstrated, and 5 indicated that it was demonstrated õto a great extent.ö These classroom observations were conducted twice during each school year, in order to decrease the likelihood that a single observation might have caught the class on an atypical day, or that it might occur on a day when there is little observable interaction taking place between the facilitator and the students (a phenomenon that is somewhat unique to EAST, due to the expectation that the facilitator õbecomes -invisibleøwhen groups are facilitating themselvesö).

Because many of the expected instructional methods and strategies are highly contextual, and because crucial aspects of the program model cannot be observed directly, cannot be inferred from only one or two observations, or are not necessarily observable in any particular class session, simple passive observation over the course of four class sessions would not by itself be sufficient to provide a fair assessment of the fidelity of program implementation. For example, while it is likely that the extent to which student projects are planned and organized by the students will be observable, it may not be possible to determine through passive observation whether the projects were selected by the students. In addition, while the EAST model indicates that student projects should normally be selected by the students, the model also accounts for circumstances when it is acceptable or even preferable to assign a project. For example, if a student is unable to identify a project even with extensive facilitation and guidance, it may be necessary to balance the benefits of a self-selected project with the harm that could be done if the student spends too much time floundering. Similarly, if a project is suggested to the students by community members or other school faculty, it is sometimes necessary for facilitators to vocally champion these ideas and offer encouragement as a tactic to foster student selection and ownership. For these reasons, the observations were followed by interviews of the facilitators that were designed to explore the background and history of the class and of any students with whom interactions were observed, and the facilitatorgs plans for follow-through with those students. This interview also provided the facilitator with an opportunity to explain the reasons for any observed interactions that appear to differ substantially from expected strategies. The appropriateness of the facilitatorsø reasons for such exceptions was then incorporated into the final observation ratings.

Prior to using the classroom observation protocol to assess fidelity at participating target schools, the instrument was pilot tested through preliminary observations that were conducted in EAST classes in non-study schools in January 2005. For this pilot-testing phase, simultaneous observations of the same class were conducted by the evaluator, the ADE Director of Technology Resources and Planning, and three EAST staff, including the National Director and two Site Health Coordinators. The observers did not communicate with each other during the observations, and each one completed the observation protocol independently. After the protocols were completed for each such joint observation, the observers compared their ratings and discussed reasons for any discrepancies; whenever possible, these debriefings took place before embarking on the next observation. The evaluator monitored the level of interrater agreement after each observation, using two different consensus estimates (after Stemler, 2004),

one based on exact agreement and one on agreement within one scale point.³⁶ While it was expected that agreement rates might be lower during the initial observations, during which the debriefing process served to calibrate observersø interpretations of the ratings with each other, the plan was to repeat the process until an interrater agreement level of at least 90% was obtained. A total of 11 EAST class sessions were observed by the group of five observers during the initial pilot process. Once the desired level of reliability was obtained, observers would conduct observations of classes in the target schools individually. At the beginning of the 2005-2006 school year, in order to ensure that observersø interrater reliability was still adequate, the first four observations of target schools conducted in October 2005 were conducted simultaneously by all three of the observers who would be responsible for the remaining observations that year (including the evaluator, the ADE Director of Technology Resources and Planning, and one of the EAST Site Health Coordinators).

For the formal observations of target classes at the eight implementing schools, two EAST sections were observed at each school, and each school (and facilitator) was observed twice during each school year. During the 2004-2005 school year, observations of Cohort 1 classes at all eight schools were conducted in February 2005 and again in April 2005; each observation was conducted by one of the five observers who had been trained during the pilot testing. In school year 2005-2006, target school observations were conducted in October 2005 and April 2006. The fall observations of Cohort 2 were conducted by three of the original five observers, including one member of the evaluation team, the EAST Director of Site Support, and the ADE Director of Technology Resources and Planning. These three individuals observed the first four classes in the fall 2005 round of observations as a group in order to determine whether their interrater agreement levels remained satisfactory. Once interrater agreement was confirmed, the evaluator scores were used for these four observations and the other two sets of observation scores were dropped from the database. Spring observations were conducted individually by one of the five observers who had participated in the original training.³⁷

Because of concerns about the potential for bias from the observers from ADE and the EAST program, efforts were made to protect the integrity of the observational data by directly addressing that potential during the training process. The first step was to develop explicit guidelines for all observers (see Appendix XIV), which all observers-in-training (including those affiliated with ADE and EAST) were required to review. This document included warningsô directed specifically towards observers affiliated with ADE and the EAST programô that it was õcrucial that they remain constantly aware of any possible bias towards believing in the effectiveness of the program, and that they remain scrupulous about not allowing any such inclinations to influence their ratings.ö In order to minimize a possible source of motivation for such bias, observers were made aware that õany tendency to provide overly favorableô or overly

³⁶ Consensus estimates provide a measure of the extent to which the different observers are in agreement about the ratings. Because the research team felt that exact agreement was an unnecessarily stringent criterion for an ordinal scale, agreement was also calculated based on the frequency with which observersøratings occurred within one scale point of each other. As noted by Stemler (2004), this is a commonly used adjustment for relaxing the criterion of exact agreement when using an ordinal scale.

³⁷ Due to unavoidable scheduling conflicts, it was not possible to have the other two project personnelô the EAST National Director and the Director of Professional Developmentô participate in the fall 2005 verification of interrater agreement. Nevertheless, these latter two observers were among the five who had initially established interrater agreement in January 2005.

unfavorableô ratings would have an unpredictable effect on the final assessment of program impact, since this ultimately depends not on absolute student performance but on studentsø progress, which can not be determined in a single observation.ö The guidelines then emphasized that õthe best way to assure an accurate assessment of the programøs impact is to provide accurate and objective ratings.ö These points were re-emphasized verbally during the training process, and the evaluator who conducted the training was convinced that the EAST observers understood and agreed with them. In order to maintain the integrity of the ratings, all observers (from Metis, ADE, and EAST) reviewed the observation protocol, annotations, and guidelines prior to each new round of observations.

Interrater Reliability. As a result of the classroom observation training preceding the winter 2005 observations, very high interrater agreement rates were obtained. Agreement was calculated based on the last six observations (following the five initial training observations). Each of the ten possible pairs of observers (five observers taken two at a time) achieved exact agreement (i.e., provided the same rating) on the overall fidelity ratings for at least four of the six observations. Averaged across all observations for all pairs, this resulted in an exact agreement rate of 77% on Overall Fidelity. Exact agreements for all ten observer pairs on each of the three Environment ratings and six sub-Environment ratings were also almost all 67% or higher. As a result, among the Environment and sub-Environment ratings, the lowest overall exact agreement rates were 73% on Educational Environment and its three sub-Environment ratings (Instruction, Class/Project Management and Nature of Projects), and on two of the three sub-Environment ratings of Environment of Expectations (Classroom Culture and Student Work); while all remaining Environment and sub-Environment ratings (Physical Environment, Environment of Expectations, and its sub-Environment of Outlook for Overall Program Achievement) averaged 80% agreement. With interrater agreement defined as ratings that differ by no more than one scale point, 100% agreement was achieved by all ten observer pairs on all ten rating categories (including overall fidelity, the three environments, and the five aspects of the environments). Indeed, even across all 11 observations (including the first five õtrainingö observations), all ten observer pairs achieved agreement within one scale point on at least 10 of 11 observations (91%) for all ten rating categories, with 100% agreement for all but five out of 100 (10 observer pairs x 10 rating categories) comparisons. These interrater agreements for all rating categories, both for exact agreements and for agreements within one point, are provided below.

Rating Scale	Exact Agreement	Agreement within One Point
I. Physical Environment	80%	100%
II. Educational Environment	73%	100%
Instruction	73%	100%
Class/Project Management	73%	100%
Nature of Projects	73%	100%
III. Environment of Expectations	80%	100%
Classroom Culture	73%	100%
Outlook for Overall Program Achievement	80%	100%
Student Work	73%	100%
Overall Fidelity Rating	77%	100%

Table 7January 2005 EAST Pilot ObservationsInterrater Reliability Based on Percent Agreement38

Agreement rates from the fall 2005 observations were, not unexpectedly, slightly lower, since it had been almost a year since the training had taken place. Nevertheless, interrater agreement among these three observers remained sufficiently high during the fall 2005 observations that it was determined that retraining was unnecessary. Specifically, exact agreement reached 83% for Physical Environment and Environment of Expectations, and while it was somewhat lower for Educational Environment and Overall Fidelity, their agreement within one scale point remained at 100% for all scales and subscales except Overall Fidelity, for which agreement was 83%. These results are shown in the following table.

³⁸ Agreement is calculated as the percent of times that a pair of observers provided the same rating (left column) or ratings within one point (right column), averaged across the last six pilot observations and across all possible pairs among all five observers.

Rating Scale	Exact Agreement	Agreement within One Point	
I. Physical Environment	83%	100%	
II. Educational Environment	67%	100%	
Instruction	67%	100%	
Class/Project Management	83%	100%	
Nature of Projects	83%	100%	
III. Environment of Expectations	83%	100%	
Classroom Culture	67%	100%	
Outlook for Overall Program Achievement	83%	100%	
Student Work	67%	100%	
Overall Fidelity Rating	67%	83%	

 Table 8

 October 2005 EAST Observations

 Interrater Reliability Based on Percent Agreement³⁹

It is well known, however, that simple percent agreement can overestimate interrater reliability, since it does not account for the amount of agreement that would be expected by chance. While this problem is less serious with scales containing more categories or levels, it nevertheless provides a more conservative assessment of reliability to use a measure that accounts for chance. One measure that is commonly used for this purpose is Cohenøs Kappa (Cohen, 1960). Although this statistic was originally developed for nominal scales and has also been noted to have its drawbacks, it is still recommended (e.g., Dewey, 1983) and is still commonly used (Bakeman, 2000). It is particularly recommended when most observations fall into a single category (or a subset of possible categories), which inflates the percent-agreement statistic (Stemler, 2004). This was the case both during the January 2005 pilot observations and during the first four observations during October 2005 that were used to re-check interrater reliability. In the pilot observations, the vast majority of ratings fell within the top three points of the scale (98% overall, compared with 60% as would be expected by chance), while in the October 2005 observations, all ratings fell between 2 and 4. While Kappa was developed for nominal scales, it provides an even more conservative estimate of reliability for ordinal scales, since it does not take into account the fact that two adjacent ratings demonstrate higher consensus than ratings at opposite ends of the scale. In order to provide the most conservative estimate to ensure that our apparently high rater reliabilities were valid, we therefore recalculated our exact agreement reliabilities using Cohenøs Kappa. The resulting Kappa values are presented in the tables below.⁴⁰

⁴⁰ Kappa was calculated using Program for Reliability Assessment with Multiple Coders (PRAM) software from Skymeg Software (2002). Retrieved on September 6, 2005, from http://www.geocities.com/skymegsoftware/pram.html

 ³⁹ Agreement is calculated as the percent of times that a pair of observers provided the same rating (left column) or ratings within one point (right column), averaged across the first four target observations and across all possible pairs among all three observers.
 ⁴⁰ Kappa was calculated using Program for Reliability Assessment with Multiple Coders (PRAM) software from

Interrater Kenability Dased on C	onen s Kappa
Rating Scale	Карра
I. Physical Environment	68%
II. Educational Environment	59%
Instruction	58%
Class/Project Management	59%
Nature of Projects	61%
III. Environment of Expectations	68%
Classroom Culture	58%
Outlook for Overall Program	680/
Achievement	08%
Student Work	63%
Overall Fidelity Rating	64%

 Table 9

 January 2005 EAST Pilot Observations

 Interrater Reliability Based on Cohen's Kappa⁴¹

Table 10
October 2005 EAST Pilot Observations
Interrater Reliability Based on Cohen's Kappa ⁴²

Interfacer Renability Dabed on Conten 5 Rappa						
Rating Scale	Карра					
I. Physical Environment	67%					
II. Educational Environment	-5%					
Instruction	33%					
Class/Project Management	33%					
Nature of Projects	70%					
III. Environment of Expectations	70%					
Classroom Culture	51%					
Outlook for Overall Program Achievement	33%					
Student Work	49%					
Overall Fidelity Rating	47%					

As the above tables show, after adjusting for chance,⁴³ average Kappa values in January 2005 ranged from 58% to 68% on all rating scales, with a reliability of 64% on Overall Fidelity. These results remain impressive, given that Kappa values above 60% may be considered õsubstantialö even for nominal scales (Landis and Koch, 1977). However, interrater reliability was more problematic during the 2005-2006 school year among the three raters who conducted observations in that year. In particular, rater reliability on the Environment of Expectations and Physical Environment, and on the Nature of Projects sub-scale of the Educational Environment, remained quite robust, at 67% to 70% above chance. However, three scales reached only õmoderateö agreement (Classroom Culture, Student Work and Overall Fidelity had Kappa values

⁴¹ Cohenøs Kappa was calculated for the last six pilot observations for all five observers.

⁴² Cohenøs Kappa was calculated for the first four observations for all three observers.

⁴³ The Kappa calculations indicate that the probability that the obtained concurrence of ratings for each scale occurred purely by chance ranged from 27% to 38%.

ranging from 47% to 51%), three demonstrated only õfairö agreement (Instruction, Class/Project Management and Outlook for Overall Program Achievement all had Kappa values of 33%), and one (Educational Environment) actually demonstrated a slightly lower percent agreement than would be expected by chance. It should be noted, however, that while a sample of observations that fall into a subset of possible categories may inflate the percent agreement statistic, it has also been shown that range restriction relative to the variance in the population results in a *reduction* in interrater reliability (*e.g.* Sackett *et al.*, 2002). Thus, considering that the Kappa estimates are overly conservative for an ordinal scale and that their values may have been further deflated by range restriction, we felt that the moderate agreement level achieved for Overall Fidelity was still satisfactory for using this measure in analyses.

<u>Construct Validity</u>. As discussed previously, a consideration reflecting on the validity of the observation protocols is the fact that not all components of the model will be rated in any given observation. This could occur if the activity that is observed in any given class session does not provide evidence relating to a particular component, or because the observer does not witness the relevant activity (or does not recognize its relevance). While the first situation is unavoidable,⁴⁴ the extent to which relevant activity is missed or misinterpreted can be minimized through training.

On the observation protocol, each component rating scale included a response option of *Don't Know*, to provide the observer with an option when no evidence relevant to a particular component was observed. However, observers were instructed that they should never leave any ratings blank. Thus, the proportion of missing ratings for each observer is relevant to the extent to which they employed the rating scales appropriately, while the extent to which the observers varied in the proportion of *Don't Know* (or missing) ratings provides further evidence of interrater reliability. In this case, however, the result of variations in the proportions of *Don't Know* or missing ratings is also relevant to construct validity, since each rating scale actually represents slightly different constructs each time it is used if the program characteristics that it is based on are not always consistent.

For the six pilot observations that the five observers conducted as a group in winter 2005, the following tables show the proportion of ratings from each observer that were scored as *Don't Know* or left blank (missing).

⁴⁴ While direct observation provides many advantages in objectivity over self-report data, this represents one of its primary weaknesses, especially when resources are not sufficient for a large enough number of observations to provide a representative time sample.

Physical Environment							
Measure	Observer	# Observations	Mean	SD	F	Sig.	
	1	6	25.0%	0.00			
	2	6	8.3%	12.91			
% Don¢t Know	3	6	16.7%	17.08	1 870	.145	
	4	6	10.4%	14.61	1.079		
	5	6	20.8%	10.21			
	Total	30	16.2%	13.19			
	1	6	0.0%	0.00			
	2	6	35.4%	32.03		367	
% Missing	3	6	12.5%	15.81	1 1 2 6		
70 1411351112	4	6	18.8%	40.12	1.120	.507	
	5	6	20.8%	39.26	1		
	Total	30	17.5%	30.01			

Table 11a January 2005 Pilot Observations Observer Consistency in % of "Don't Know" and Missing Ratings Physical Environment

Table 11bJanuary 2005 Pilot ObservationsObserver Consistency in % of "Don't Know" and Missing RatingsEducational Environment

Measure	Observer	# Observations	Mean	SD	F	Sig.
% Donøt Know	1	6	19.1%	10.59		
	2	6	3.6%	6.08		
	3	6	8.8%	5.20	7 201	000
	4	6	3.0%	4.69	7.301	.000
	5	6	16.1%	3.89		
	Total	30	10.1%	8.95		
	1	6	1.2%	2.20	26.012	
	2	6	50.0%	12.21		
% Missing	3	6	5.5%	4.60		.000
% Missing	4	6	12.1%	16.33		
	5	6	5.5%	4.60		
	Total	30	14.8%	20.30		

Environment of Expectations						
Measure	Observer	# Observations	Mean	SD	F	Sig.
	1	6	13.5%	8.20		
	2	6	7.1%	15.28		
% Donøt Know	3	6	13.5%	11.44	1.056	122
	4	6	1.6%	3.80	1.930	.152
	5	6	16.7%	10.75		
	Total	30	10.5%	11.28		
	1	6	3.2%	3.89		
	2	6	44.4%	11.13		
% Missing	3	6	6.3%	5.77	26.027	.000
70 Wilssing	4	6	5.6%	13.61	20.027	
	5	6	2.4%	3.98		
	Total	30	12.4%	18.23		

Table 11cJanuary 2005 Pilot ObservationsObserver Consistency in % of "Don't Know" and Missing RatingsEnvironment of Expectations

Table 11d
January 2005 Pilot Observations
Observer Consistency in % of "Don't Know" and Missing Ratings
Overall Fidelity

Measure	Observer	# Observations	Mean	SD	F	Sig.
	1	6	18.3%	8.90		
	2	6	5.0%	5.55		
% Dongt Know	3	6	10.7%	6.73	6 6 1 5	001
% Dolla Klow	4	6	3.4%	5.02	0.015	.001
	5	6	16.7%	4.76		
	Total	30	10.8%	8.49		
	1	6	1.6%	1.79		
	2	6	47.2%	8.90		
% Missing	3	6	6.3%	2.22	25.273	.000
	4	6	11.1%	17.36		
	5	6	6.2%	4.73		
	Total	30	14.5%	18.91		

As the above data show, the proportion of program components within Physical Environment, averaged across all six pilot observations, that were rated as *Don't Know* by each observer ranged from 8% to 25%, from 3% to 19% within Educational Environment, from 2% to 17% within Environment of Expectations, and from 3% to 18% for Overall Fidelity. These differences among observers were significant⁴⁵ for the Educational Environment ratings, on which Observers 1 and 5 had larger proportions of *Don't Know* ratings than any of the other three, and on Overall Fidelity, on which Observers 1, 3, and 5 rated more items as *Don't Know* than Observers 2 or 4.

Similarly, the proportion of program components for which the ratings were missing for each observer, within each of the three environments and the overall rating averaged across all six observations, ranged from 0% to 35%, 1% to 50%, and 2% to 44%, and 2% to 47%,

⁴⁵ Based on one-way Analysis of Variance (ANOVA).

respectively. These differences among observers were statistically significant for all ratings except Physical Environment. In all other ratings, Observer 2 clearly had the largest proportions of missing ratings.

The proportions of ratings from each observer that were scored as *Don't Know* or left blank during the fall 2005 group observations of the second cohort of EAST classes are summarized below.

Table 12aOctober 2005 Group ObservationsObserver Consistency in % of "Don't Know" and Missing RatingsPhysical Environment

Measure	Observer	# Observations	Mean	SD	F	Sig.	
% Dongt Know	1	4	15.6%	11.97			
	2	4	0.0%	0.00	13.364	.002	
	5	4	25.0%	0.00			
	Total	12	13.5%	12.45			
% Missing	1	4	0.0%	0.00			
	2	4	6.2%	7.22	3 000	.100	
	5	4	0.0%	0.00	5.000		
	Total	12	2.1%	4.86			

Table 12b October 2005 Group Observations Observer Consistency in % of "Don't Know" and Missing Ratings Educational Environment

Measure	Observer	# Observations	Mean	SD	F	Sig.
	1	4	13.6%	9.21		
% Dong Know	2	4	10.0%	4.33	1.348	.308
% Dong Know	5	4	17.7%	5.43		
	Total	12	13.8%	6.87		
% Missing	1	4	1.4%	1.74		
	2	4	1.8%	2.10	.290	.755
	5	4	0.9%	1.05		
	Total	12	1.4%	1.57		

October 2005 Group Observations Observer Consistency in % of "Don't Know" and Missing Ratings									
	En	vironment of Exp	oectations						
Measure Observer # Observations Mean SD F Sig.									
	1	4	3.6%	4.56					
0/ David Warner	2	4	2.4%	2.75	.391	.687			
70 Donga Khow	5	4	4.8%	3.89					
	Total	12	3.6%	3.59					
% Missing	1	4	0.0%	0.00					
	2	4	0.0%	0.00					
	5	4	0.0%	0.00					
	Total	12	0.0%	0.00					

Table 12c

Table 12d **October 2005 Group Observations Observer Consistency in % of "Don't Know" and Missing Ratings Overall Fidelity**

Measure	Observer	# Observations	Mean	SD	F	Sig.
	1	4	11.3%	8.10		
% Dong Know	2	4	7.1%	2.92	2.054	.184
% Dong Know	5	4	15.2%	4.50		
	Total	12	11.2%	6.12		
% Missing	1	4	0.9%	1.14		
	2	4	1.8%	2.06	765	403
	5	4	0.6%	0.69	.705	.495
	Total	12	1.1%	1.39		

As these tables show, the proportion of ratings that were left blankô an error in scoring procedure that was discussed among the observers after the January 2005 observationsô declined substantially, at least during pilot observations among these three observers, in October 2005. In addition, as a result of this correction, there were no significant differences among observers in the proportion of missing ratings. While there was a significantly lower proportion of Don't Know ratings on Physical Environment from Observer 2, this might have been true in part because this observer left items blank rather than scoring them as Don't Know. This raises a question as to whether Observer 2 was interpreting the observation scale constructs differently on these items.

While these differences among observers raise concerns about interrater reliability and construct validity of the individual subscales, the problem appears to have been less severe during the Cohort 2 observations, which were used for the final outcome analyses. In addition, the stronger reliability of the final fidelity ratings, as discussed in the previous section, provide reassurance that the broader program characteristics on which each observer was basing his or her ratings were nevertheless comparable within each observed session. However, even where differences between observers were minimal, the relatively high proportions of missing or Don't Know ratings which occurred in some cases might mean that the program characteristics (constructs) on which each session was rated were not always consistent from session to session.

Discriminant validity. Related to the question of whether the observation ratings are reflecting the same construct for each observed session and for each observer, also of interest is whether each of the scales and subscales of the observation protocol are in fact measuring unique constructs, or whether the ratings for each subscale mostly reflect an overall fidelity construct. Pearson correlations among pairs of scores were conducted to determine the extent to which each scale varied independently of the others.

Results of the scale correlations (presented in Appendix XV) demonstrate that the Physical Environment scale was the most clearly distinct from other scales and subscales.⁴⁶ Lending support to the instrumentøs discriminant validity, correlations among subscales from different environments were generally somewhat lower than those among subscales from the same environment. This distinction was not strong, however, and all scales and subscales (other than Physical Environment) were closely related to each other.⁴⁷

While these results indicate that the value of the observations for measuring discreet aspects of program fidelity may be limited, they also provide additional reassurance that, despite the number of missing and Don't Know ratings, the constructs that are being rated are similar from observer to observer.

End-of-Year Program Ratings

Given the complexity of the facilitator-student dynamic and the large number of elements specified in the logic model that constitute program fidelity, it seems very likely that individual class observations may not provide opportunities to rate all aspects of program fidelity. Indeed, this was underscored by the substantial proportions of missing and Don't Know ratings on the observation forms. For this reason, additional program ratings were obtained in order to supplement the formal observations with additional sources of information about program fidelity.

In addition to the individual class observations, a number of other program-related activities take place throughout the year that provide the national EAST staff with opportunities to become more familiar with the EAST program and facilitator at each school. These include periodic site health visits to EAST classrooms, training activities (for students, facilitators, and building administrators) and regional conferences, as well as ongoing correspondence and communications by email, telephone, and listserv. An end-of-year rating form was created (see Appendix XVI) that enabled EAST staff to provide an overall rating of program fidelity for each of the eight experimental schools, based on the insights gained through these additional interactions. This form was structured the same as the observation protocol, but was used to reflect all information about each instructor program implementation that the EAST staff could glean from their interactions with the facilitator throughout the entire year.

In order to standardize this process, the facilitators at each school participating in the study all received the same number of site health visits during each year of the outcomes study. Ratings were completed via consensus among the three EAST staff membersô including the

⁴⁶ Pearson correlations between the Physical Environment scale and all other scales ranged from .58 to .74; all correlations were statistically significant. ⁴⁷ Pearson correlations for all other scale pairs were above .80; all were statistically significant.

National Director, the Director of Professional Development and the Director of Site Support who conducted site visits during the year and who had the greatest contact with the school-based EAST staff through the above activities. All three of these EAST staff were among the observers who were trained for the original classroom observations. As was the case for the formal observation ratings, EAST staff were trained to base their end-of-year ratings on observable indicators of the facilitatorøs understanding of program philosophy, strategies, instructional methods, and response to studentsøprogress, but not on the extent to which students were reaching their goals. The importance of accurate ratings and the unpredictable nature of the effects of ratings bias on evaluation outcomes were re-emphasized in order to help insure objectivity in the ratings.

Student Survey

Several additional outcomes relating to student perceptions were measured less formally through a student survey that was developed collaboratively by EAST national staff, ADE staff, and the evaluator. Student surveys (with slightly different versions administered to EAST and control students, as discussed belowô see Appendix XVII) were administered in spring 2005 and spring 2006. Because this survey was not formally pilot tested and validated, and because, in order to keep the paperwork burden on the students to a minimum, it was administered on a post-only basis, it provides a less rigorous measure of program effects than the other instruments. Nevertheless, it does possess strong face validity and therefore provides valuable suggestive information about several important program characteristics. The topics covered on the student surveys included the following:

• Teachersø/FacilitatorsøInstructional Style

The approach which EAST facilitators take to instruction represents an important interim objective of the program and, as such, an important aspect of program fidelity. While the Fidelity Observation Protocol included several sections relevant to this outcome (Rating II.A. Instruction; Rating II.B. Classroom Management; Subscale III.A. Classroom Culture), since the dynamic between the facilitator and student is such an important aspect of these outcomes, we felt that it would be instructive to ask the students themselves about their perceptions of their teachers and facilitators. The EAST students were asked the extent to which they agreed with statements about their facilitator (1=strongly disagree, 5=strongly agree) such as õMy EAST facilitator help(s) me learn how to solve my own problems in my class work,ö õMy EAST facilitator help(s) me feel comfortable working with other students who are different than me,ö and õMy EAST facilitator encourages me to do challenging class work or projects.ö In order to assess whether EAST facilitators were perceived differently from other teachers, control students were asked the same questions, but were asked to agree or disagree about these statements in reference to õmost of my teachers.ö⁴⁸

• Technology skills

⁴⁸ In order to keep the survey length reasonable, EAST students were not asked these questions about their own non-EAST teachers; however, such comparisons were discussed during the focus group interviews.

Although the EAST model considers technology a tool rather than an end in itself, familiarity with the technological tools with which EAST students are equipped is important to the success of many types of projects and is integral to other student outcomes in the model, such as õStudents develop solutions to community-based problems using emerging technology within the context of service learning projects.ö In order to determine whether EAST students had more exposure to various kinds of technologies and developed more skill in using them, the survey asked students to report their perceptions of their familiarity and skills with various common technologies (which most middle and high school students would be expected to encounter in the present day school environment), and with more specialized technologies which are representative of the tools which are provided through the EAST program and are probably less commonly encountered in classes outside of EAST.

• Self-directed learning style

õSelf-directed, student-centered learningö is one of the four pillars of the EAST model, while õresponsibility for self-directed learningö is one of the primary objectives of the modeløs environment of expectations. The student survey tapped into possible differences between EAST and control studentsø learning styles in the context of completing school projects by asking respondents whether they agreed with statements about issues such as the locus of responsibility for making sure that they learn and their teachersø/facilitatorøs role in helping them complete a project.

• Plans for after high school

Finally, the EAST program is believed to build studentsø motivation and confidence, including their inclination to õreflect on their abilities and set new goals for continuous personal improvement and contributions to society.ö This objective was assessed, in part, through survey questions about how much education EAST and control students expected to complete and how they expected to spend their time after leaving high school.

Supplemental Recruitment Survey

Because student participation in EAST is in almost all cases voluntary, and because the processes for recruitment and screening of students are locally defined and vary widely by school, the possibility arises that these processes could create substantial selection effects that would confound the study findings. However, not all processes or policies that result in narrowing the field of EAST participants necessarily cause a selection effect. Selective recruitment and screening that limit participation in the program to the population for which it was intended would not cause a selection bias, although it would be important to limit generalizations about the impact of the program on that population. For example, while EAST is designed to develop characteristics such as ability to collaborate and willingness to work hard, program leadership believes that students must at least enter the program with a willingness to become hard-working collaborators in order to succeed in EAST.

However, recruitment, screening, and selection policies that change the demographics of the population in other ways would run a considerable risk of introducing bias. We compared the demographic characteristics of EAST students to non-EAST students in the same schools in order to determine whether EAST students were representative of the population, or whether any differences from the school population were consistent with the population towards which the model is targeted. To a large extent, however, the characteristics that the model defines as determining a studentø appropriateness for the program, while detectable by teachers who know the student well, are not limited to simple demographics or test scores that are easily documented. Since it would have been prohibitively time consuming and intrusive to attempt to empirically measure all of the salient characteristicsô which include hard-to-measure traits such as ability to work in an unstructured environment and comfort in a hands-on settingô it was determined that the best approach would be to document each schoolø recruitment and screening policies and infer whether these policies could be expected to result in EAST classes that were consistent with the intended population. This was accomplished by comparing the target schoolsø policies to those that are recommended by the EAST model.

The recruitment survey (Appendix XVIII) was administered to the principal and facilitator of each of the eight target EAST schools, to explore what characteristics were considered by the schools in the process of recruitment, screening, and/or placement. For each of 26 character traits that might be expected to increase a studento chance of success in the program, respondents were asked to indicate how important they felt the trait was to success in EAST; whether their school attempted to obtain students who possess the indicated characteristic, attempted to obtain diversity among EAST students in terms of the characteristic, or did not consider the trait as a factor during the recruitment, screening and selection process at their school; and, for characteristics they reported their school attempted to obtain in EAST candidates, how much of that characteristic a candidate for the EAST program should possess, according to their schooløs policy. Similarly, for each of 11 character traits that might hinder a studentøs success in EAST, respondents were asked to indicate how important it was to a studentøs success in EAST that they not possess the characteristic; whether their school attempted to obtain students who possess the indicated characteristic, attempted to obtain diversity among EAST students in terms of the characteristic, or did not consider the trait as a factor during the recruitment, screening, and selection process at their school; and, for characteristics they reported that their school attempted to obtain in EAST candidates, how much of that characteristic a candidate for the EAST program should possess, according to their schooløs policy.⁴⁹ Finally, the survey asked whether each school attempted to obtain students from specific demographic groups, or attempted to obtain diversity in terms of particular demographic characteristics. While it is clear that the schools may not always be successful in obtaining a group of students with the characteristics that they seek, at the same time it is also certainly less likely that there is any systematic bias in that population if school policies actively seek to avoid such bias.

A comparison of the consistency between principalsø and facilitatorsø responses to these questions was used as a partial indicator, for formative evaluation purposes, of the quality of communications and program management efforts at each site. The extent to which schoolsø selection and screening criteria were found to be consistent with the model were included as an additional measure of fidelity of program implementation for the final outcome analyses. Where substantive differences were found between facilitatorsø and principalsø descriptions of their

⁴⁹ These last two ratings may seem counterintuitive in the context of hindering characteristics, but may be relevant for some schools that may offer the EAST program as an alternative for academically struggling students.

school policies, comparisons to the model were based on the responses from the individual who was reported (in the original principal and facilitator surveys) to make the final decision about placing students into EAST classes.

Student Inventories

An exhaustive search of the literature was conducted in order to identify published and previously validated instruments that could be used to assess some of the student characteristics and habits of mind (as discussed under Purpose of the Research above, and depicted in the Logic Model in Appendix I) that EAST is designed to develop and that are believed to contribute to academic and career success. These areas were identified as those skills that EAST, Inc. and ADE staff consider to be among those that are most directly influenced by participation in EAST. The skills of greatest interest, on which the instrument search focused, included critical thinking and life skills such as motivation, teamwork, cooperation, and problem solving. The objective of the search was to identify instruments designed to measure the skills that had demonstrated at least minimum standards of reliability and construct validity, had the greatest face validity for measuring the skills that the EAST program strives to build, and had demonstrated sensitivity to instructional interventions (as opposed to measuring primarily developmental traits). Because the assessments were to be administered on a pre-post (fall-spring) basis, in order to minimize testing effects, it was also hoped that instruments could be found that had at least two alternate but comparable forms.

As a result of this search, two published and validated student inventories were identified that assess two important characteristics that are influenced by EAST.⁵⁰ The identified instruments included the Inventory of School Motivation (ISM), developed by Dennis M. McInerney (McInerney, 2004; McInerney et al., 2001; McInerney, 1998; McInerney et al., 1997), and the Social Problem Solving Inventory for Adolescents (SPSI-A), developed by Marianne Frauenknecht and David R. Black (1995; 2003). In order to minimize the addition to studentsø already very full testing schedule, the instruments were modified slightly (with permission of the authors) to allow them to be administered in one sitting. For the ISM, the modification involved selecting five subscales from the numerous options in the item bank that the author had developed. These subscales were selected as having the strongest face validity for the objectives of the program and having demonstrated high reliability from past research. For the SPSI-A, only those items comprising the Problem Solving Skills scale were used. Because the SPSI-A was not being administered in whole, and because the research team was interested in looking at effects on individual subscales, the long version of the instrument, which has higher reliability at the subscale level, was used. These excerpts utilized the most relevant portions of each instrument while still making it possible to administer both instruments in a single one-hour session. Since the original inventories were not designed to be administered alongside each other, the possibility that the experience of completing one inventory might have an impact on responses to the other inventory was checked by randomizing the order in which the ISM and SPSI-A were administered to each student. No significant differences were identified in the

⁵⁰ While these are certainly not the only traits that the program develops, nor are they necessarily the most important ones, it was simply impractical to attempt to measure all of the immediate outcomes identified in the program modelô although several other outcomes, including self-directed learning style, career aspirations, and familiarity and skills with technology, were measured less formally on the student survey, discussed below.

inventory results based on testing order for either the treatment or control students, nor for any subpopulations.

Validity and Reliability. Development and validation studies of the original instruments conducted by the authors demonstrated strong psychometric properties for both the SPSI-A and ISM. For the SPSI-A, content validation was achieved through screening for face validity by reviewers with expertise in social problem solving, item reduction through review of correlations between items and scale scores, and simplification of reading level compared to the adult version of the instrument from which it was derived. Additional item analyses were conducted by evaluating reliability and inter-item correlations, and removing or replacing weakly correlated items. For the scales used in the EAST study (the Problem Solving Skills Scale and its subscales), the resulting instrument demonstrated internal reliabilities ranging from .79 to .95, and stability coefficients over a two-week period ranging from .63 to .77. Construct validity was further demonstrated through intracorrelations with another problem-solving measure, which correlated with the SPSI-A Problem Solving Scale at r = .73, and through concurrence with a measure of personal problems which showed that a person with higher problem-solving abilities (as measured on the SPSI-A) reported fewer personal problems. Details of instrument development for the SPSI-A can be found in Frauenknecht and Black (1995) and are summarized in the table below.

The ISM was designed as an exploratory instrument that provides an extensive item bank of questions relevant to motivation in cross-cultural contexts. Individual items can be selected for use according to need, but the authors also provide several scales that were created through *a priori* alignment of content and validated through exploratory and confirmatory factor analyses, which have been conducted in several studies in diverse populations, especially focusing on atrisk academic performance. These studies have developed a variety of scales constructed through various combinations and permutations of the original items.

A summary of the characteristics of the original ISM and SPSI-A instruments is presented in the table below. 51

⁵¹ Because of copyright restrictions for the SPSI-A, the Student Inventory form can not be publicly distributed and is not shown in the appendices. Copies of the original instruments can be obtained from the authors.

					Reliability	
	Populations	Age		Domains/Subscales	Test-Retest	Internal
Scale	Tested	Appropriateness	Readability	Used for EAST Study	(Pearson's r)	(Cronbach's α) ⁵²
SPSI-A	1,062 students, Midwest	Recommended for	õFog indexö	Problem-solving Skills	.77 (2 wks)	.9495
	metro and rural high	adolescents	ranged from grade	Problem Identification	.65	.7992
	schools, mostly white		equivalent of 6.8	Alternative Generation	.70	.8589
		Pilot test mostly 9 th &	to 9.0	Consequence Prediction	.63	.7890
		10 th graders		Implementation/Evaluation	.72 ⁵³	.8489 ¹
				Reorganization	NA	.87
ISM	Total of 9,731 students	Test populations	NA	Effort	.79 (1 yr)	.7079
	across numerous studies	ranging from grade 7-		Social Concern	.70 (õ õ)	.6572
	of Anglo, migrant and	12		Praise	NA	.7284
	aboriginal Australian,			General Mastery	NA	.71
	Navajo, American,			General Social	NA	.74
	Nepalese, African,					
	Montagnais and Chinese					
	students					

Table 13 Summary of Characteristics of the Original ISM and SPSI-A Instruments

Sources: Frauenknecht & Black, 1995; McInerney, 2004; McInerney et al., 2001; McInerney, 1998; McInerney et al., 1997.

 ⁵² Values for ISM reflect American (including Anglo and Navajo), African and Chinese students and across all groups.
 ⁵³ In some assessments, Implementation/Evaluation and Reorganization were treated as a single subscale.

Because the demographics of the student population in the EAST study were considerably different from those of many of the populations that were used in the original development of the ISM and SPSI-A; because, in the case of the ISM, different studies often used variations of the motivation scales that were based on slightly different combinations of items; and because the SPSI-A was not administered in its entirety (only one of the three major scales was used) and the instruments were administered in tandem, it was felt that the psychometric properties of these instruments should be confirmed in the context in which they were used for this study. Post hoc analyses of the EAST Student Inventory results were therefore conducted to determine whether the scales hung together as originally defined, and whether the instrument demonstrated acceptable reliability.

In order to verify the integrity of the scales, confirmatory factor analyses⁵⁴ were conducted on the responses from the fall 2005 Student Inventories. Separate analyses were conducted on the ISM items and the SPSI-A items. These analyses showed that responses from students in the EAST study were highly consistent with the original scale definitions, with items õbehavingö as scales almost exactly as specified by the authors. The only exceptions were as follows:

<u>SPSI-A</u>:

- The items from the Problem Identification (PID) and Alternative Generation (ALT) scales behaved as a single scale.
- One item that was originally defined as part of the PID scale (õWhen I have a problem, I examine the things that surround me which may cause the problemö) was associated more strongly with the Consequence Prediction (CON) scale.
- The items from the CON scale and most of the items from the Evaluation (EVAL) scale behaved as a single scale.
- One of the items from the original EVAL scale (õAfter carrying out a solution to the problem, I decide what went right and what went wrongö) was associated more strongly with the Reorganization (REO) items, and another (õI often solve my problems and achieve my goalsö) was associated more strongly with the Implementation (IMP) items.

<u>ISM</u>:

- One of the items from the original Effort (EFF) scale (õI don¢t mind working a long time at schoolwork that I find interestingö) was not clearly associated with any scale.
- One of the items from the original General Mastery (GMA) scale (õI am most motivated when I am solving problemsö) was not clearly associated with any scale.
- Two of the items from the original General Social (GSO) scale were more strongly associated with the Social Concern Scale.

Detailed results from the factor analyses are presented in Appendix XIX.

⁵⁴ Using Maximum Likelihood method with Varimax rotation and Kaiser normalization.

In addition to the factor analyses, the student inventories were also analyzed for internal consistency reliability. These analyses were also conducted separately for the ISM items and the SPSI-A items.⁵⁵ Results of these reliability analyses are presented in the tables below.

Table 14
Student Inventory—ISM Scales
Internal Reliability Analyses

Administration	Seeles	Cronbach's a				
Aummstration	Scales	All students	Control	Experimental		
	Effort	.825	.832	.813		
	Praise	.845	.856	.833		
Fall 2004	Social Concern	.703	.728	.680		
	General Mastery	.764	.749	.774		
	General Social	.727	.755	.699		
	Effort	.835	.811	.848		
	Praise	.809	.809	.806		
Spring 2005	Social Concern	.732	.704	.750		
	General Mastery	.809	.795	.814		
	General Social	.762	.740	.774		

Table 15 Student Inventory—SPSI-A Scales Internal Reliability Analyses

Administration	Seeles		Cronbach's α	
Administration	Scales	All students	Control	Experimental
	Problem Identification	.781	.797	.762
	Alternative Generation	.816	.839	.792
	Consequence Prediction	.786	.793	.780
Fall 2004	Implementation	.778	.785	.771
	Evaluation	.778	.790	.763
	Reorganization	.844	.850	.838
	Problem-solving Skills	.945	.949	.942
	Problem Identification	.868	.858	.875
	Alternative Generation	.874	.856	.888
	Consequence Prediction	.850	.840	.859
Spring 2005	Implementation	.798	.765	.824
	Evaluation	.846	.835	.850
	Reorganization	.863	.853	.870
	Problem-solving Skills	.964	.960	.967

As the above data show, the scales of the ISM and SPSI-A that were used for the student inventories demonstrated internal reliability among control students as well as EAST participants that was quite comparable to the results reported by the authors. Across all students, alpha values of the SPSI-A scales on the fall 2004 and spring 2005 administrations ranged from.78 to

⁵⁵ These analyses were conducted only on the Cohort 1 results.

.96, and for the IMS scales, from .70 to .84, indicating a degree of reliability that was clearly adequate for the research.

One additional consideration relevant to the validity of the Student Inventories was brought to the evaluatorsø attention during the spring 2006 student focus groups. During this interview, several students stated that they believed that some of their peers were getting frustrated with the redundancy of completing the same inventory twice a year and what they perceived as redundancy in the questions, and had started responding randomly or making arbitrary patterns on the answer form. While the consistency of the post hoc validation analyses with the authorsø original instrument development provided considerable reassurance that most students were not responding arbitrarily, the potential for such a response to invalidate the study results was serious enough to warrant additional examination of the data.

In order to determine the extent to which this might have occurred, the fall 2005 inventories were examined for visual and statistical evidence of random or arbitrary responses.

The first step involved conducting statistical analyses to identify inconsistencies in a given student¢s responses (e.g., one õagreeö and one õdisagreeö response) to pairs of items which were believed to address the same or similar constructs.⁵⁶ While a substantial proportion of respondents (33.3%) had at least one or two discrepant pairs on the SPSI-A, only a handful (1.3%) had more than three (out of a total of 13 identified pairs among the 32 inventory items). On the ISM, only 18% of respondents had at least one discrepant pair, and only 0.5% had more than two (out of a total of 8 identified pairs among the 27 inventory items). It is perhaps not surprising that a small percentage of students might have an occasional set of responses that appear contradictory, perhaps due to fatigue or misreading a question, and we found it encouraging that only a very small proportion of respondents had significant numbers of inconsistent responses.

While these results were encouraging, we recognized that if students were drawing geometric patterns on their answer keys (some of the focus group participants referred to students õChristmas tree-ingö their responses), these pairs analyses might not detect it, since similar items on these inventories were scaled in the same direction and tended to occur in close proximity (especially on the SPSI-A). For this reason, all of the fall 2005 answer forms were also inspected visually. Visual inspection entailed looking at the original answer forms to see if there appeared to be any regular geometric pattern in the response choices. Based on the visual inspection, approximately 3% of the ISM and 4% of the SPSI-A inventories had what we considered to be flagrantly regular response patterns that we inferred to be arbitrary.

⁵⁶ It should be noted that these pairs were identified by the evaluator, not derived from lie scales created by the publishers. Consider for example the following two items from the SPSI-A:

a) õI decide if a problem is part of a larger, more complex problem that should be solved first.ö

b) õWhen I have a problem, I find out if it is part of a bigger problem that I should deal with.ö

While there are subtle construct differences between these two statements that a respondent may or may not discern, we believe that they are similar enough that we expected that a student providing thoughtful self-ratings would be very unlikely to describe one of them as being õtrue of meö and the other as õnot at all true of me.ö

While the results of these visual inspections raised somewhat greater concern than the item pair analyses, the overall proportion of flagged surveys remained reasonably low. Given the strong results from other evidence of instrument validity and reliability discussed above, it was concluded that the likelihood that these response patterns might bias study results in a meaningful way seemed negligible.

Iowa Tests

While specific academic skills such as reading and mathematics are not explicitly targeted as immediate outcomes of the EAST program, increased motivation and the broader life skills such as problem solving and critical thinking which *are* explicitly targeted are also believed to play an important role in academic achievement (as well as promoting success in other areas of life), as discussed in the Research Context section of this report. For this reason it was decided to monitor the programøs impact on primary academic indicators as well as the immediate outcomes assessed by the student inventories and surveys. The Iowa Tests (Iowa Test of Basic Skills for grades 6-8 and the Iowa Test of Educational Development for the high school grades [ITBS/ITED]), developed at the University of Iowa College of Education and published by Riverside Publishing, Inc., were selected for this purpose. Aside from the practical consideration that these tests were already being used in Arkansas as the state achievement tests, they offered several advantages that made them particularly appropriate for this study.

- the Iowa Tests include levels that are specifically designed for a range of grades that is inclusive of the grades participating in the EAST study (6-12);
- test results include a developmental standard score (DSS) that is vertically scaled across all grade levels of the ITBS and ITED;
- each level has alternate forms, making it possible to administer fall-spring pre-post testing within the same year while minimizing the possibility of confounding the results due to testing effects; and
- while some achievement tests generate scores that incorporate mechanical skills such as math computation and vocabulary, the core batteries of the ITBS/ITED generate scores that separate these from more conceptual skills.

The ITBS and ITED are standardized, norm-referenced tests that include different literacy and mathematics sections that address higher order thinking skills, interpretation, classification, comparison, analysis and inference. Among their intended purposes, the instruments include providing achievement information that makes it possible to monitor year-to-year developmental differences. The ITBS Reading Comprehension test consists of passages including fiction, fables, tales, poetry, interviews, diaries, biographical sketches, science and social studies materials, and other nonfiction. Approximately two-thirds of the questions require students to draw inferences or to generalize about what they have read. The ITED Reading Comprehension score provides information about the kinds of comprehension skills students are expected to continue to develop as they proceed through high school.

The ITBS and ITED mathematics batteries include scales in concepts and estimation, problem solving and data interpretation, an optional computation scale, and a total score. EAST program staff determined that this content is well aligned with the broader goals of the EAST
program, and the developersøstatement of intended purposes supports the use of the tests for the evaluation.

Further evidence of the validity of the ITBS is based on subtest intercorrelations with the Cognitive Abilities Test, predictive association with future grades and test performance, studies of cognitive problem solving processes used to complete the test, and equating studies related to score meaning among others, was found to be very positive (Brookhart, 1998). The ITED presents support for criterion and construct-related validity through evidence such as disattenuated correlations with the listening assessment (.68-.79 with the ITED composite), correlations between the complete battery and constructed response supplements (.60s to low .70s), and correlations with the other tests including the CogAT, and Work Keys, as well as for the American College Testing program (ACT), for which it was the model (Mehrens, 1998).

Utilizing the Kuder-Richardson Formula 20 (KR20), the ITBS was found to be highly reliable. For levels 12 through 14 of the exam (grades 6 through 8), the ITBS had KR20 values ranging from .90 to .92 on the Reading *Comprehension* portion, and from .92 to .94 on the Math Total for *Problems and Data Interpretation* (excluding *Computation*) portion (Hoover, Dunbar, and Frisbie, 2001, pp. 24-26). KR20 values for the ITED (grades 9-12) subtests are almost all in the .80s or .90s.

The ITBS and ITED scores used for the EAST study included, from the mathematics batteries, the concepts and problem solving scores (excluding computation), and for the reading battery, the reading comprehension scores (excluding vocabulary). These particular skills are better aligned with the goals of the EAST program than the more mechanical skills of computation and vocabulary.

Facilitator and Student Focus Group Interviews

The same focus group interviews that were conducted in the first year of the implementation study and were held in all three years also served to function (in the second and third years) to inform the outcomes study. During the two years of the outcomes study, participants in the focus groups were selected from among the eight EAST study schools rather than from the population of all EAST programs in Arkansas. While this may have made the interviews somewhat less generalizable to the state as a whole, it made the findings more directly relevant to the outcomes study. Participants included the facilitators from all eight implementation schools in a single interview. For the student focus groups, one to two students were selected from each school by their facilitator, and separate interviews were conducted with middle school and high school students. The content of the focus groups was summarized in the discussion of the implementation study above.

Procedures

The table below summarizes the various data sources used and how the data obtained from each source related to the studyøs research questions. Data from program documentation, student surveys, student inventories, and the Iowa Tests were collected both for EAST and control students, while program-specific data sources (EAST classroom observations, end of year program ratings, on-line principal and facilitator surveys, and supplemental recruitment surveys) were obtained for the eight implementing EAST schools.

	Data Sources									
	(Annual Administration Schedules)									
Research Questions	EAST classroom observations (Fall, Spring)	End of year program ratings (Spring)	Online principal survey (Spring)	Online facilitator survey (Spring)	Supplemental Recruitment Survey (Spring)	School and program documentation (Ongoing)	Student survey (Spring)	School Motivation/Problem Solving Inventories (Fall, Spring)	Iowa Tests (Fall, Spring)	
Student attitudes			\checkmark	\checkmark			\checkmark	\checkmark		
Student skills and content knowledge			\checkmark	\checkmark			\checkmark	\checkmark	\checkmark	
School support and culture			\checkmark	\checkmark	\checkmark		\checkmark			
Facilitatorsøattitudes	\checkmark	\checkmark	\checkmark	\checkmark			\checkmark			
Facilitatorsøpractices	\checkmark	\checkmark		\checkmark	\checkmark		\checkmark			
Program fidelity/implementation conditions	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark					
Other factors affecting outcomes (student demographics, locale, studentsø# years in EAST, other exposure to learning, student training, etc.)						\checkmark	\checkmark		\checkmark	

Table 16Overview of East Outcome Study

Although facilitator attitudes and practices are among the targeted outcomes of the EAST program, these factors are intermediate outcomes that are targeted by the program for the purpose of improving student outcomes. In the context of this study, facilitator attitudes and practices were defined as components of overall program fidelity and treated as intermediate explanatory variables that might influence the extent to which the EAST program affects studentsøattitudes and skills. Another way of viewing this is that the EAST program may not be expected to accomplish what its developers expect unless it is being implemented correctly.

The study methods and analyses were designed to determine the extent to which the EAST program influenced each of the following outcomes, after controlling for the effects of other possible factors. The specific outcome variables that were explored in this studyô from among those outcomes believed to be influenced by the EAST program (as outlined in the logic model, Appendix I)ô and the instruments or data sources that were used to measure these variables are summarized in the table below.

Variable	Measurement Instrument/Source
Academic performance ó reading	Spring ITBS/ITED: Reading Comprehension Score
	Spring ITBS Math Total Score (excluding Computation)/ITED Math
Academic performance ó mathematics	Concepts & Problems Score
Motivated to work hard	Spring Student Inventory: ISM ó Effort score
Motivated by external praise ⁵⁷	Spring Student Inventory: ISM ó Praise score
Motivated to help others	Spring Student Inventory: ISM ó Social Concern score
Motivated by accomplishment/achieving	
success	Spring Student Inventory: ISM ó General Mastery score
Motivated by working in groups	Spring Student Inventory: ISM ó General Social score
Problem solving skills	Spring Student Inventory: SPSI-A ó Problem Solving Skills score
Skill at defining the characteristics of a problem	Spring Student Inventory: SPSI-A ó Problem Identification score
Skill at identifying alternative solutions	Spring Student Inventory: SPSI-A ó Alternative Generation score
Skill at predicting the outcomes of a solution	Spring Student Inventory: SPSI-A ó Consequence Prediction score
Skill at creating strategies for implementing a solution	Spring Student Inventory: SPSI-A ó Implementation score
Skill at assessing the outcomes of a solution	Spring Student Inventory: SPSI-A ó Evaluation score
Skill at revising strategies in response to the assessment of outcomes	Spring Student Inventory: SPSI-A ó Reorganization score
Self-directed learning style	Spring Student Survey: Learning Style composite score
Motivation to pursue further learning	Spring Student Survey: Plans for After HS composite score

Table 17EAST Outcome Study:Outcome Variables and How They Were Measured

A number of õexplanatoryö factors were explored that might have an impact on any of the above outcomes. These included characteristics of the student as well as characteristics of their school or classroom setting, and in both cases also included characteristics of the studentøs experience (or absence thereof) in the EAST environment as well as factors that are not directly related to the EAST program. These explanatory variables and the instruments or data sources that were used to measure them are summarized in the tables below.

⁵⁷ It would be expected that the EAST program would *reduce* the extent to which students are motivated by external praise.

Attribute	Measurement Instrument/Source
Grade level	ADE records
English proficiency	ADE records (ELL Y/N)
Gender	ADE records
Reading proficiency	Fall 2005 ITBS/ITED Reading Comprehension
Math proficiency	Fall 2005 ITBS Math Total ⁵⁸ /ITED Concepts & Problems
Race	ADE records (coded as Minority Y/N^{59})
Socioeconomic status	ADE records (free/reduced lunch eligibility Y/N)
Special education status	ADE records (Y/N)
Gifted status	ADE records (Y/N)
Prior academic performance ó reading	Fall ITBS/ITED: Reading Comprehension score
	Fall ITBS Math Total score (excluding Computation)/ITED Math
Prior academic performance ó mathematics	Concepts & Problems score
Prior motivation to work hard	Fall Student Inventory: ISM ó Effort score
Prior motivated by external praise	Fall Student Inventory: ISM ó Praise score
Prior motivation to help others	Fall Student Inventory: ISM ó Social Concern score
Prior motivated by accomplishment/improvement	Fall Student Inventory: ISM ó General Mastery score
Prior motivated by group work	Fall Student Inventory: ISM ó General Social score
Prior problem solving skills	Fall Student Inventory: SPSI-A ó Problem Solving Skills score
Prior tendency to define the characteristics of a	Fall Student Inventory: SPSI-A ó Problem Identification score
problem	
Prior tendency to identify alternative solutions	Fall Student Inventory: SPSI-A ó Alternative Generation score
Prior tendency to predict the outcomes of a	Fall Student Inventory: SPSI-A ó Consequence Prediction score
solution	
Prior tendency to create strategies for	Fall Student Inventory: SPSI-A ó Implementation score
implementing a solution	
Prior tendency to assess the outcomes of a	Fall Student Inventory: SPSI-A ó Evaluation score
solution	
Prior tendency to revise strategies in response to	Fall Student Inventory: SPSI-A ó Reorganization score
the assessment of outcomes	
Cohort (# of years in EAST)	ADE records (coded as 0 for control students)
# regular math, English, technology or science	ADE records
courses enrolled during SY05-06 (separate count	
# anriched meth English or soloned courses	ADE records
# children main, English of science courses enrolled during SV05-06 (separate court for each	ADE ICUIUS
entoned during 5105-00 (separate count for each	

Table 18a **EAST Outcome Study:** Student Level Explanatory Variables

 ⁵⁸ Excluding the Computation section.
 ⁵⁹ Original records included the following ethnic categories: Asian, Black, Hispanic, Native American, White/non-Hispanic

School and Program Level Explanatory variables							
Attribute	Measurement Instrument/Source						
School/Cla	ssroom Level Variables						
School locale	NCES records (õruralö/õurbanö ⁶⁰)						
School level	ADE records (middle school, high school, middle/high school) ⁶¹						
Total school enrollment	ADE records						
Program status	Study records						
Program fidelity	Fidelity Observations: Total Fidelity score (control schools=1) ⁶²						

Table 18b EAST Outcome Study: School and Program Level Explanatory Variables

The influence of these explanatory variables was explored for each of the above outcome variables separately, using Hierarchical Linear Modeling (HLM). These analyses were conducted as follows.

A separate model of best fit was built for each outcome variable, through an iterative model building process. First, Level 1 (student level) predictor variables were entered into the model. For the ITBS outcomes, ITBS reading and math pretest scores, student demographic characteristics (grade, gender, minority status, ELL status, special education status and gifted status), course enrollment count variables and pre-test scores from the school motivation student inventory subscales and from the total problem solving inventory were all entered into the initial model. In addition to these predictors, several interaction terms that seemed intuitively likely to be relevant to each outcome were also included in the initial full model. These included gender by special education status; and free/reduced lunch eligibility by each of the eight pre-scores (including 5 ISM scores, total problem solving, and ITBS reading and math).⁶³ From this initial model, those predictors that were least significant (had the highest *p*-value and/or lowest coefficient value) were removed from the model one by one until only predictors with significant p-values remained.

For the Student Inventory outcomes (problem solving and school motivation), an analogous approach was used to build the Level 1 model, except that variables were entered in groups, in order to accommodate limitations of the software. The order in which these groups were entered was based on conceptual likelihood of predictorsørelation to the outcome variable. For these models, all of the Inventory pretest scores (including subscale scores on the problem

⁶⁰ As defined above for the implementation study target schools.

⁶¹ All EAST programs in the study were administered only at the middle or the high school level even among several middle/high schools in the target group. However, since the statistical methods used for analysis require that all cases (including both implementing and control schools) be coded on all variables, Program Level could not be included as an explanatory variable but was replaced by school level and student grade.

⁶² Because each facilitator received multiple observation ratings in each year (usually four), a single fidelity score was created for each facilitator in each year by averaging all observations for the year. This is explained further in the Results section below.

⁶³ Due to limitations of the HLM software, in some cases it was not possible to include all variables in the initial full model. In these cases, the inclusion of interaction terms and counts of enriched courses were prioritized based on the likelihood that they would be relevant to a particular outcome. (For example, interactions involving ISM pretest scores were more likely to be relevant to the ISM outcomes than to the SPSI-A or ITBS outcomes. Counts of enriched courses were given the lowest priority since they occurred infrequently in the sample.)

solving inventory for problem solving outcomes), ITBS pretest scores and student demographics were entered initially. Variables were extracted from the model one by one until only predictors with significant p-values remained. Then, the interaction terms mentioned above were added to the model. Again, those predictors that were least significant (had the highest *p*-value and/or lowest coefficient value) were removed from the model until model of best-fit was obtained that included only predictors with significant p-values.

Once the most parsimonious model of Level 1 predictors was obtained, an exploratory analysis of Level 2 variables on the Level 1 predictors was conducted. Along with estimates of the magnitude of each Level 1 predictors influence on the outcome, output also includes estimates of the variance in these coefficients (i.e., how well the coefficient represents the influence of that predictor for all students in the sample). An exploratory analysis offers suggestion into which Level 2 (school level) characteristics are likely to explain away any significant variance. Level 2 variables were entered into the model as suggested by this output, until the final modelô containing both Level 1 and/or Level 2 predictors, as appropriateô was obtained. P-values for both coefficients (i.e., estimates of the magnitude of influence) and variances (i.e., how homogenous the predictorsø influence is across all individuals in the data) were considered until this best-fitting model was finalized.

In addition to the formal statistical modeling, many of these variables and their relationships to EAST outcomes were further explored through additional data collection procedures, including individual and focus group interviews with facilitators, students, school and district administrative staff, and ADE and EAST, Inc. staff; school visits; and additional measures, including the End-of-Year Program Ratings, supplemental recruitment surveys, and additional items on the student surveys. Findings from these supplemental sources are discussed in the relevant sections below.

Results

Fidelity of Program Implementation

EAST Classroom Observations

The primary measure used to assess the programøs impact on facilitatorsø attitudes and instructional practices were the classroom observations. Within each program year, the four sets of scores generated for each facilitator (two observations in fall/winter and two in spring) were averaged to create a single composite score. These composite scores were then averaged across the eight facilitators, or (for analyses by program level or locale) across facilitators within level or locale. This approach was used rather than computing a simple average of all observations for all facilitators for several reasons. First, in one case (in spring 2005), one of the facilitators was observed three times instead of twice. By collapsing observations into a single composite score for each facilitator, we ensured that this individualøs scores were not given more weight in the final analyses. In addition, each observation, which lasted no more than an hour, would obviously not be representative of the many hours of EAST class that take place during the year. While the four observations still may not be statistically representative of the entire year, at least they provide more opportunities to observe more of the activities and interactions that do not occur in every class session, as well as some opportunity to capture some of the seasonal

differences that may exist between beginning and end-of-year activities. For this reason, the composite scores were considered to provide a better overall assessment of a facilitatorøs approach, and to provide a more accurate sense of the range of fidelity among facilitators.

Additional measures that were used to assess the programøs impact on facilitatorsø fidelity to the model included the end-of-year program ratings and on-line principal and facilitator surveys. As explained in the Measures/Instruments section above, the end-of-year ratings were completed by consensus by the team of EAST, Inc. personnel at the end of each program year. These ratings were compared to the composite observation ratings for each facilitator to verify whether the two measures seemed to confirm or contradict each other. Finally, facilitator and principal surveys and the facilitator and student interviews provided qualitative data to lend greater insight into the meaning of the rating results.

Year Two EAST Classroom Observations

A total of 33 classes were observed (winter 2004 and spring 2005) in the eight implementing schools, including four observations (two in winter and two in spring) of seven of the eight facilitators, and five observations (two in winter and three in spring) of the eighth.⁶⁴ The following table summarizes the demographics of the observed classes.

	Program	m Level	Program		
Characteristics of students	Middle School	High School	Rural	Urban	Total
Total N of classes observed	17	16	25	8	33
Average N of students per class	13.4	8.8	10.7	12.6	11.2
% Female	57%	38%	48%	53%	50%
% White non-Hispanic	53%	84%	76%	36%	65%

Table 19SY 2004-2005 Classroom Observation FormsCharacteristics of Participating Classes and Students

As the above table shows, demographic distributions were different in the observed middle school classes than in those in the high schools. There were fairly equal distributions of girls and boys, and of minorities and non-minorities, in the middle school classes, while in the high school classes, there were considerably fewer girls than boys and substantially more white students than minorities. By program locale, classes were evenly distributed between girls and boys in observed classes in both the rural and urban settings; however, minority students formed a substantial majority of urban classes but were in the minority in the rural classes. All observed classes were small, averaging around 11 students.

Results of the fidelity observation ratings are summarized in the tables that follow. The first table presents overall fidelity ratings from the first implementation year (2004-2005).

⁶⁴ Because one of the observed classes was interrupted by a visit by school administrators, the observer decided to conduct an additional observation with that facilitator in case the first observation was invalidated. However, since the evaluator and EAST staff concluded that the first observation should still be considered validô especially since it is inherent in the program that EAST students have little established routine and are accustom to modifications to what routine might existô both observations were included in the analysis.

Patings	Total N of	Moon	Ra	nge	#	#
Katings	Facilitators	cilitators Wean		Max.	DK ⁶⁵	Missing ¹
I. Physical Environment	8	4.4	3.5	5.0	0	0
II. Educational Environment	8	3.3	1.5	4.8	0	0
A. Instruction	8	3.3	2.0	4.5	0	0
B. Class/Project Management	8	3.3	1.5	4.8	0	0
C. Nature of Projects	8	3.3	1.5	4.3	0	0
III. Environment of Expectations	8	3.3	1.5	4.5	0	0
A. Classroom Culture	8	3.2	1.0	4.5	0	1
B. Outlook for Overall Program Achievement	8	3.3	1.0	5.0	2	0
C. Student Work	8	3.4	1.5	4.5	0	0
OVERALL FIDELITY	8	3.3	1.5	4.5	0	0

Table 20 SY 2004-2005 Classroom Observation Forms Summary of Ratings – All Classes

Mean fidelity averaged across all eight facilitators indicated that the physical classroom setup in these programs was highly appropriate to an EAST classroom, with an average physical environment rating of 4.4 on a 5-point scale and a range from 3.5 to 5.0. An example of a classroom rated as a 5.0 in physical environment was one that had all EAST required technology available and in working order, as well as comfortable rolling chairs for students and seating arrangements that facilitated collaboration, with ample space for small group work.

Following are some examples of class sessions that received higher and lower ratings in the Educational Environment and Environment of Expectations, designed to provide a clearer picture of what these ratings represent. These ratings are derived from actual observations during the two years of the outcomes study.

An example of a class that was rated as a 4 for Educational Environment was one in which students got right to work on projects and the facilitator helped get students started by directing them towards a project that was due to a client shortly. The facilitator walked around the room asking questions, giving suggestions, and encouraging students when needed, while at the same time leaving students and groups to themselves when they were operating smoothly. Students were clearly aware of the organization and expectations of the class. Projects were interesting and used a variety of both sophisticated and basic technologies and met the needs of the school and larger community, ranging from a GPS mapping of trees in the area, to collecting and refurbishing computers for underprivileged children, to removing red-eye from prom pictures for the yearbook.

In contrast, in a classroom rated as a 2 in Educational Environment, only about half of the students got right to work, while others talked and checked their personal email. Although one student began working with the plotter printer to replace a fallen poster, most students spent their time gossiping and doing little productive work, while the facilitator made little active effort to coach students into productive activities. The limited number of projects were not õEAST-likeö (for example, they were not student led or organized) and several students talked about finding a

⁶⁵ Numbers of *Don't Know* and missing responses are based on individual observations, not composites.

new project with little success. Many students spent the time learning new technologies with no goal in mind. There was very little facilitation to help students find good projects; rather, the facilitatorøs input was to tell students that if they couldnøt find something useful to do they would just have to do technology tutorials.

An example of a class that was rated as a 4 on Environment of Expectations was a classroom where the facilitator worked with students, walked around the room, and helped guide their work while assisting them to think through ideas themselves. The class was flexible, yet students clearly knew what to expect from the facilitator and what was expected from them. Group work was the norm in this classroom, although some students worked independently for specific purposes. The classroom culture was open but organized, and students were clearly comfortable asking questions of the facilitator (who often pointed them to other resources) and other students in the lab.

In a class that received a rating of 2 on the Environment of Expectations, students were often off-task, with little consequence, and they were not engaged in õrealö community-based projects that were guided by students working in groups. The classroom was quiet as students worked largely on their own and didn¢t interact much with one another. While students seemed to be left to their own devices, the facilitator had no format for managing student activities or monitoring their progress. Students seemed to be õkilling timeö in the class with no particular purpose, and their workô for example, making PowerPoint presentations for no one in particularô focused on product over purpose, process, and student development. There seemed to be low expectations from students, who spent much of their period sitting and wasting time. The facilitator did not take advantage of õteachable momentsö with students or guide students toward rich, useful, collaborative work.

During the first year of program implementation, the eight study schools were, on average, already well along in establishing EAST programs that followed the national model, although with mean fidelity ratings of 3.3 for the Educational Environment and Environment of Expectation scales and for Overall Fidelity, there was still considerable room for improvement. In addition, there was a wide range of implementation fidelity in these areas, with facilitators ranging from 1.5 to 4.8 in their Educational Environment ratings, and from 1.5 to 4.5 in both their Environment of Expectations and Overall Fidelity ratings.

Comparisons of fidelity ratings for EAST programs by level (implemented at the middle school vs. high school grades) and by locale (urban vs. rural) are summarized in the following tables.

	Program	Total		Ran		#	#
Ratings	Level	N of Facilitators	Mean	Min.	Max.	DK	Missing ⁶⁶
I Physical Environment	MS	4	4.3	3.5	5.0	0	0
1. Thysical Environment	HS	4	4.5	4.0	5.0	0	0
II Educational Environment	MS	4	3.3	1.5	4.8	0	0
11. Educational Environment	HS	4	3.2	2.5	4.0	0	0
A Instruction	MS	4	3.4	2.0	4.5	0	0
	HS	4	3.2	2.3	4.0	0	0
B Class/Project Management	MS	4	3.4	1.5	4.8	0	0
D. Class/110jeet Management	HS	4	3.3	2.3	4.3	0	0
C Nature of Projects	MS	4	3.1	1.5	4.3	0	0
C. Nature of Projects	HS	4	3.5	3.0	4.0	0	0
III Environment of Expectations	MS	4	3.3	1.5	4.5	0	0
III. Environment of Expectations	HS	4	3.3	2.8	4.0	0	0
A Classroom Culture	MS	4	3.2	1.0	4.5	0	0
A. Classicolli Culture	HS	4	3.3	2.8	4.0	0	1
B Outlook for Overall Program Achievement	MS	4	3.1	1.0	5.0	2	0
B. Outlook for Overall Hograni Aemevement	HS	4	3.5	2.8	4.5	0	0
C. Student Work	MS	4	3.3	1.5	4.5	0	0
	HS	4	3.5	3.0	4.0	0	0
OVERALL FIDELITY	MS	4	3.3	1.5	4.5	0	0
UTERALL FIDELITI	HS	4	3.3	2.8	4.0	0	0

Table 21SY 2004-2005 Classroom Observation FormsSummary of Ratings by Program Level

⁶⁶ Numbers of *Don't Know* and missing responses are based on individual observations, not composites.

Datings	Program	Total N of	Maan	Ra	nge	#	#
Katings	Level	Facilitators	Mean	Minimum	Maximum	DK	Missing ⁶⁷
I Physical Environment	Rural	6	4.4	4.0	5.0	0	0
1. Thysical Environment	Urban	2	4.3	3.5	5.0	0	0
II Educational Environment	Rural	6	3.0	1.5	4.0	0	0
II. Educational Environment	Urban	2	3.9	3.0	4.8	0	0
A Instruction	Rural	6	3.1	2.0	4.0	0	0
A. Instruction	Urban	2	3.9	3.3	4.5	0	0
B Class/Project Management	Rural	6	3.1	1.5	4.3	0	0
D. Class/110ject Management	Urban	2	4.0	3.3	4.8	0	0
C Nature of Projects	Rural	6	3.3	1.5	4.0	0	0
C. Nature of Projects	Urban	2	3.5	2.8	4.3	0	0
III. Environment of	Rural	6	3.1	1.5	4.0	0	0
Expectations	Urban	2	3.8	3.0	4.5	0	0
A Classroom Culture	Rural	6	3.0	1.0	4.3	0	1
A. Classroom Culture	Urban	2	3.8	3.0	4.5	0	0
B. Outlook for Overall Program	Rural	6	3.1	1.0	4.5	2	0
Achievement	Urban	2	4.0	3.0	5.0	0	0
C. Student Work	Rural	6	3.3	1.5	4.3	0	0
	Urban	2	3.6	2.8	4.5	0	0
OVEDALL FIDELITY	Rural	6	3.1	1.5	4.0	0	0
OVERALL FIDELITT	Urban	2	3.9	3.3	4.5	0	0

Table 22 SY 2004-2005 Classroom Observation Forms Summary of Ratings by Locale

While there were no notable differences in mean fidelity among schools implementing EAST at the middle vs. high school levels, it is interesting to note that the middle school programs consistently showed a wider range of fidelity, with lower minimum ratings and higher maximum ratings than the high schools on all ten ratings except physical environment, where the highest ratings for both groups reached the top of the scale.

On the other hand, schools operating in an urban environment implemented the model with greater fidelity, on average, than the rural schools, particularly within the Educational Environment and Environment of Expectations. Overall Fidelity ratings averaged 3.9 for urban schools, compared to 3.1 for schools in a rural setting, while Educational Environment ratings averaged 3.9 and 3.0 for urban and rural schools, respectively, and 3.8 and 3.1, respectively, for Environment of Expectations. Other than Physical Environment, urban schools demonstrated the greatest fidelity in Outlook for Overall Program Achievement and Class/Project Management, with mean ratings of 4.0 in both areas, while the rural schools demonstrated the greatest fidelity in Nature of Projects and Student Work, with mean ratings of 3.3 in each area.

Year Three EAST Classroom Observations

In the third and final year of the evaluation a total of 32 classes were observed, again with four observations (two in fall 2005 and two in spring 2006) conducted of each facilitator, with

⁶⁷ Numbers of *Don't Know* and missing responses are based on individual observations, not composites.

the context of observed classes distributed the same as the prior year. Demographics of the observed classes are summarized below.

Characteristics of Participating Students											
	Program	m Level	Program								
Characteristics of students	Middle	High	Rural	Urban	Total						
	School	School	Kurai	Orban							
Total N of classes observed	16	16	24	8	32						
Average N of students per class	13.3	12.4	12.9	12.6	12.8						
% Female	45%	48%	54%	25%	47%						
% White non-Hispanic	44%	89%	74%	37%	66 %						

Table 23SY 2005-2006 Classroom Observation Forms
Characteristics of Participating Students

Demographic distributions of students in the observed classes were again different for middle vs. high schools and for rural vs. urban schools. While there were fairly equal distributions of girls and boys in both middle and high school classes, the urban classes had a much lower representation of girls (25% on average). As in the previous year, distribution of minority and non-minority students was again fairly even in the middle schools, but high schools and rural schools consisted predominantly of white students, while in urban schools only about one third of the student population, on average, was white. All observed classes were again small, albeit slightly larger than the prior year, averaging around 13 students.

As in year two of the evaluation, the ratings from these observations were collapsed within schools to provide a single set of ratings for each of the eight schools. Overall fidelity ratings from the second implementation year (2005-2006) are summarized in the following table.

Summary of Ratings – All Classes										
Ratings	Total N of	Moon	Ra	nge	#	#				
Turing 5	Facilitators	acilitators Nicali	Minimum	Maximum	DK	Missing ⁶⁸				
I. Physical Environment	8	4.1	2.5	5.0	0	0				
II. Educational Environment	8	3.5	1.5	5.0	0	0				
A. Instruction	8	3.5	1.5	4.8	0	0				
B. Class/Project Management	8	3.5	1.8	5.0	0	0				
C. Nature of Projects	8	3.3	1.0	4.5	0	0				
III. Environment of Expectations	8	3.4	1.5	4.5	0	0				
A. Classroom Culture	8	3.4	2.0	4.5	0	0				
B. Outlook for Overall Program	8	35	15	5.0	0	0				
Achievement	0	5.5	1.5	5.0	0	0				
C. Student Work	8	3.3	1.5	4.5	0	0				
OVERALL FIDELITY	8	3.5	1.5	5.0	0	0				

 Table 24

 SY 2005-2006 Classroom Observation Forms

 Summary of Patings

In their second year of program implementation, the eight target schools demonstrated a slight increase in fidelity in the EAST model from the first year of implementation in all ratings. This is consistent with the improvements that one would hope the facilitators would achieve with

⁶⁸ Numbers of *Don't Know* and missing responses are based on individual observations, not composites.

an additional year to become comfortable with the program and an additional year of training. The mean fidelity ratings for Overall Fidelity and for Educational Environment both were 3.5 (compared with 3.3 in the prior year), with ranges from 1.5 to 5.0, while the Environment of Expectations ratings averaged 3.4 (compared with 3.3 in the prior year), with a range from 1.5 to 4.5. Although there was slight improvement, these eight schools continued to show quite a wide range of levels of implementation. The only exception to the improvement from the prior year was in the physical setup of the labs, which, while still rating high (the mean rating for Physical Environment was 4.1), was slightly lower than the previous year (4.4). The low end of the range of ratings for the physical environment (from 2.5 to 5.0) was also lower than the prior year, when the low end of the range was 3.5 on the 5-point scale.

Comparisons of fidelity ratings for EAST programs by level (implemented at the middle school vs. high school grades) and by locale (urban vs. rural) are summarized in the following tables.

		01 10000080	~				
Datinga	Program	Total N of	Moon	Ra	nge	#	#
Katings	Level	Facilitators	wiean	Minimum	Maximum	DK	Missing ⁶⁹
I Physical Environment	MS	4	3.8	2.5	5.0	0	0
1. Fllysicai Environment	HS	4	4.4	4.0	4.5	0	0
II Educational Environment	MS	4	3.4	1.5	5.0	0	0
II. Educational Environment	HS	4	3.6	3.3	4.0	0	0
A Instruction	MS	4	3.4	1.5	4.8	0	0
A. Instruction	HS	4	3.6	3.3	4.0	0	0
B. Class/Project	MS	4	3.4	1.8	5.0	0	0
Management	HS	4	3.6	3.3	4.0	0	0
C Natura of Projects	MS	4	3.2	1.0	4.5	0	0
C. Nature of Projects	HS	4	3.4	3.0	3.5	0	0
III. Environment of	MS	4	3.3	1.5	4.5	0	0
Expectations	HS	4	3.6	3.0	4.0	0	0
A Classroom Culture	MS	4	3.4	2.0	4.5	0	0
A. Classiooni Culture	HS	4	3.4	3.0	4.0	0	0
B. Outlook for Overall	MS	4	3.4	1.5	5.0	0	0
Program Achievement	HS	4	3.6	3.3	4.0	0	0
C. Student Work	MS	4	3.2	1.5	4.5	0	0
C. Student WOIK	HS	4	3.5	3.0	4.0	0	0
	MS	4	3.4	1.5	5.0	0	0
OVERALL FIDELITY	HS	4	3.6	3.3	4.0	0	0

Table 25 SY 2005-2006 Classroom Observation Forms Summary of Ratings by Program Level

⁶⁹ Numbers of *Don't Know* and missing responses are based on individual observations, not composites.

Datings	Program	Total N of	Moon	Ra	nge	#	#
Katings	Level	Facilitators	Mean	Minimum	Maximum	DK	Missing ⁷⁰
I Drysical Environment	Rural	6	4.3	3.0	5.0	0	0
1. Thysical Environment	Urban	2	3.5	2.5	4.5	0	0
II Educational Environment	Rural	6	3.5	1.5	5.0	0	0
II. Educational Environment	Urban	2	3.6	3.3	4.0	0	0
A Instruction	Rural	6	3.4	1.5	4.8	0	0
A. Instruction	Urban	2	3.8	3.5	4.0	0	0
B. Class/Project	Rural	6	3.5	1.8	5.0	0	0
Management	Urban	2	3.5	3.0	4.0	0	0
C Nature of Projects	Rural	6	3.2	1.0	4.5	0	0
e. Nature of Hojeets	Urban	2	3.6	3.3	4.0	0	0
III. Environment of	Rural	6	3.4	1.5	4.5	0	0
Expectations	Urban	2	3.5	3.0	4.0	0	0
A Classroom Culture	Rural	6	3.4	2.0	4.5	0	0
A. Classiooni Culture	Urban	2	3.6	3.3	4.0	0	0
B. Outlook for Overall	Rural	6	3.5	1.5	5.0	0	0
Program Achievement	Urban	2	3.5	3.0	4.0	0	0
C Student Work	Rural	6	3.3	1.5	4.5	0	0
C. Student Work	Urban	2	3.4	2.8	4.0	0	0
	Rural	6	3.5	1.5	5.0	0	0
UVEKALL FIDELITY	Urban	2	3.5	3.0	4.0	0	0

Table 26 SY 2005-2006 Classroom Observation Forms Summary of Ratings by Locale

While the differences were not dramatic, there were some notable distinctions in fidelity between middle and high schools and, to a lesser extent, between rural and urban EAST schools. High schools and rural schools tended to have a better physical setup in their EAST classes: in Physical Environment, middle schools received a mean rating of 3.8 as compared to high schools, which received a mean rating of 4.4. Similarly, rural schools scored a mean rating of 4.3 and urban schools scored a lower mean rating of 3.5. In addition, while mean ratings were quite similar in the other Environments and in Overall Fidelity, the amount of variation among schools was considerably greater for middle schools than for high schools in almost all areas.⁷¹

End-of-Year Program Ratings

The End-of-Year Program Ratings were based on the same form that was used for the classroom observations. While the latter ratings were retrospective and relied more on memory than the observation ratings, they had the advantage over the program ratings in that they were based on a full year of interactions (including additional classroom visits, telephone and email conversations, and activities during Partnership Conferences among others) between each schooløs EAST program and the entire team of EAST, Inc. staff. As such, these ratings reflected a much more representative cross section of program activities and characteristics, and EAST staff were more easily able to address a number of factors that could not be assessed through passive observation alone. Since the EAST staff completed these forms through consensus, only

⁷⁰ Numbers of *Don't Know* and missing responses are based on individual observations, not composites.

⁷¹ The amount of variation was also greater among rural schools than urban schools, but this may simply reflect the fact that only two of the eight study schools were in an urban locale.

one set of ratings was generated for each program in each year. Results of these ratings are summarized below.

Table 27
SY 2004-2005 End-of-Year Program Rating Forms
Summary of Ratings – All Schools

Detinge	Total N of	Moon	Range		
Ratings	Facilitators	Mean	Minimum	Maximum	
I. Physical Environment	8	4.38	3	5	
II. Educational Environment	8	3.38	2	4	
A. Instruction	8	4.00	3	5	
B. Class/Project Management	8	3.38	2	4	
C. Nature of Projects	8	3.88	2	5	
III. Environment of Expectations	8	3.63	2	5	
A. Classroom Culture	8	3.25	1	5	
B. Outlook for Overall Program Achievement	8	3.75	2	5	
C. Student Work	8	3.63	2	4	
OVERALL FIDELITY	8	3.88	3	5	

Table 28SY 2005-2006 End-of-Year Program Rating FormsSummary of Ratings – All Schools

Patings	Total N of	Moon	Range		
Katings	Facilitators	Witali	Minimum	Maximum	
I. Physical Environment	8	4.25	2	5	
II. Educational Environment	8	3.63	2	5	
A. Instruction	8	3.75	2	5	
B. Class/Project Management	8	3.63	2	5	
C. Nature of Projects	8	3.50	1	5	
III. Environment of Expectations	8	3.63	2	5	
A. Classroom Culture	8	3.63	2	5	
B. Outlook for Overall Program Achievement	8	3.63	2	5	
C. Student Work	8	3.63	2	5	
OVERALL FIDELITY	8	3.63	2	5	

By comparing the results of these year-end ratings to those of the observation ratings discussed in the previous section, we can determine whether the two measures tend to converge. Discrepancies between the two measures might exist because they are measuring different constructs (this possibility was the reason that the year-end ratings were implemented in the first place), or because one or the other (or both) of the measures are not reflecting the intended constructs accurately, or for a combination of these reasons. However, if the measures converge, it provides reassurance of the construct validity of both measures, since it is unlikely that poor construct validity would result in the same sorts of errors in ratings on two different instruments.⁷²

⁷² This is only true to the extent that the instruments use different approaches. Since the year-end ratings were based on the format of the observation protocol, it is possible that the ratings could converge even if problems in construct validity did exist within dimensions on which the two instruments overlap.

A comparison of the two rating sources reveals that there were some non-trivial discrepancies on the 2004-2005 ratings (compare Tables 20 and 27 above). The most substantial discrepancies occurred in two of the sub-ratings within Educational Environment, one of the sub-ratings within Environment of Expectations, andô most notablyô on the Overall Fidelity ratings, in which the differences between the mean ratings ranged from 0.4 to 0.6 scale points. On the 2005-2006 ratings, however, the two measures converged considerably: none of the mean ratings or sub-ratings differed between the two measures by more than 0.25 scale points, and the mean Overall Fidelity ratings were virtually identical, differing by only 0.13 scale points (see Tables 24 and 28). For the 2005-2006 ratings, these similarities held up quite well even when comparing the ratings for individual facilitators. Since the final analyses used only the 2005-2006 fidelity ratings as a control variable, it was determined that it was unnecessary to create composite fidelity ratings from both instruments, and the original observation ratings were used instead.

Recruitment and screening procedures

An important additional aspect of program fidelity that could not be addressed through classroom observations is the approach that EAST schools take to advertising the program at the school, recruiting students for participation, and screening and placing interested students into EAST classes. Since these factors were not reflected in the fidelity ratings, and since the number of explanatory variables to be included in the statistical models was threatening to become unwieldy, it was unrealistic to incorporate recruitment and screening procedures into the statistical model. Instead, it is necessary to interpret the findings of program impact within the context of the hypothetical implications of differences in recruitment and screening procedures.

It should be pointed out that the EAST model itself does not advocate that the program is appropriate for everyone, and to the extent that any pre-screening of participants is consistent with the model, it should not be regarded as creating a bias in the conclusions of the study; rather, these specifications would indicate that any conclusions about program effectiveness should only be generalized to the types of students for whom the program was intended. On the other hand, if the types of students who participate in the program are filtered (whether actively or passively) in ways that diverge from the intent of the model, it could create selection effects that threaten the generalizability of the study s conclusions for the intended populations.⁷³ It should be further pointed out, however, that what is best for the program, and the students, is not always consistent with what is best for an easily-interpretable studyô and the statege and districtsøpriorities must of course always lie with the former. In this case, while it is the ultimate goal of the EAST model to obtain diverse, representative groups of student participants, this may not be a top priority in the early years when the programs are first becoming established. Indeed, there may be advantages to first creating a successful program with a somewhat more selective group of students, and expanding the program to serve increasing numbers of more at-risk students as it becomes established. In the current study, the use of newly-established programs represented an unavoidable trade-off between the advantage of starting with a oclean slate,o versus the disadvantages (some of which are also discussed in other sections of this report) of attempting to measure impact with programs that were less mature and in some ways less consistently representative of the model.

⁷³ A demographic comparison of target EAST students and non-EAST students in the target EAST schools is discussed later in this section.

During both the 2005 and 2006 focus groups, facilitators from the study schools confirmed their perceptions of the value of EAST for a broad range of students. They observed that those who were the most successful in traditional classes are not necessarily the most successful in EAST, which does not provide the same structure, while others whose academics had been impeded by shyness, antisocial behavior, lack of maturity, or lack of aptitude in traditional learning modalities have flourished in EAST success unique environment. The facilitators also reported that disabled students often do well in the program, although it can sometimes be difficult to fully accommodate their needs, especially if the class is too large or if there are too many students with special needs.

While facilitators believed in the value of the program for a broad range of students, they also indicated during interviews that there are other types of students for whom EAST may be less appropriateô including, for example, students with serious behavioral or attitude problems, as well as õgrade seekersö and students who are accustomed to greater structure and more formal measures of progress. (As a result, students who are otherwise very successful academically may not necessarily be well-suited for EAST.) It is difficult, however, to ascertain whether the limitations that facilitators perceived went beyond the expectations of the model.

In an effort to obtain a more definitive picture, principals and facilitators were surveyed about their schoolsørecruitment and screening procedures. These surveys were administered to all eight EAST study schools in spring 2005 and again in spring 2006. Since earlier interviews and surveys conducted for the implementation study indicated that these two groups sometimes had different impressions of school policies surrounding the EAST program, similar questions were posed to both groups, with the expectation that agreement between the two would tend to verify the accuracy of reported policies, while any disagreements would need to be examined for their implications about actual practices. The facilitators from all eight schools in spring 2005 and from seven of the eight schools in spring 2006. Summaries of the recruitment and selection methods that respondents reported that their school was using are presented in tables 27 and 28 below, while further analyses of whether the principal and facilitator from the same school agreed with each other are presented in tables 29 and 30.

	Principa	l Surveys	Facilitator Surveys		
Recruitment Methods	Spring 2005	Spring 2006	Spring 2005	Spring 2006	
Letter to parents	5	5	2	2	
School-wide announcements	7	6	3	4	
Posters and bulletin boards	6	6	2	6	
School newspaper or newsletter	4	3	3	4	
Advertised at feeder school(s)	1	1	2	1	
Recommended by classroom teachers	6	6	4	5	
Recommended by principal or guidance counselor	7	6	6	3	
Career Days/Career Orientation	1	1	3	4	
Demonstrations/presentations by EAST students	4	6	2	4	
Parent meetings	4	5	1	1	
Open house for incoming grades	6	5	2	2	
õPeer recruitmentö (word of mouth)	7	5	6	7	
Facilitator recruitment	7	6	6	8	
Participation is required of certain students	3	1	0	0	
Total N Responding	8	7	8	8	

Table 29Principal and Facilitator SurveysStudent Recruitment Strategies

Table 30Principal and Facilitator SurveysStudent Selection Procedures

Solation Proceedures	Prin Sur	cipal veys	Facilitator Surveys		
Selection riocedures	Spring 2005	Spring 2006	Spring 2005	Spring 2006	
First come, first served	4	5	1	2	
Random selection	1	4	1	3	
Written application	5	0	1	2	
Interviewed by facilitator or other faculty	4	3	2	3	
Efforts that reflect the range of academic abilities in the school	4	7	3	2	
Efforts that reflect the ethnic makeup of the school	4	7	4	1	
Preference given to students who have participated in EAST in previous years	4	6	3	5	
Preference given to students based on request of parents, teachers, school counselors	4	5	4	4	
Preference given to resource, special needs, or low performing students	1	1	2	2	
Preference given to students for other reasons	0	0	0	0	
Other	0	0	0	1	
None of the above (open to all students)	2	0	1	2	
Don't know			2	0	
Total N Responding	8	7	7	8	

Table 31
Principal and Facilitator Surveys
Agreement Between Principals and Facilitators About Recruitment Strategies

	2005 Surveys						2006 Surveys			
Recruitment Strategies		# Sch Princi Facilitato	iools: ipal & or Agreed	# Sc Princ Facilitato	hools: cipal & r Disagreed		# Scho Princip Facilitator	ools: oal & Agreed	# Sc Princ Facilitato	hools: cipal & r Disagreed
	Valid N	Used	Not Used	Principal Said it was Used	Facilitator Said it was Used	Valid N	Used	Not Used	Principal Said it was Used	Facilitator Said it was Used
Letter to parents	8	2	3	3	0	7	2	2	3	0
School-wide announcements	8	3	1	4	0	7	3	1	3	0
Posters and bulletin boards	8	2	2	4	0	7	5	1	1	0
School newspaper or newsletter	8	2	3	2	1	7	2	2	1	2
Advertised at feeder school(s)	8	1	6	0	1	7	1	6	0	0
Recommended by classroom teachers	8	4	2	2	0	7	4	0	2	1
Recommended to particular students by principal or guidance counselor	8	5	0	2	1	7	3	1	3	0
Career Days/Career Orientation	8	0	4	1	3	7	1	3	0	3
Demonstrations/presentations by EAST students	8	1	3	3	1	7	4	1	2	0
Parent meetings	8	0	3	4	1	7	1	2	4	0
Open house for incoming grades	8	2	2	4	0	7	2	2	3	0
õPeer recruitmentö (word of mouth)	8	5	0	2	1	7	5	0	0	2
Facilitator recruitment	8	5	0	2	1	7	6	0	0	1
Participation is required of certain students	8	0	5	3	0	7	0	6	1	0

Table 32
Principal and Facilitator Surveys
Agreement Between Principals and Facilitators About Selection Procedures

2005 Surveys							2006 Surveys			
Selection Procedures		# Schools: Principal & Facilitator Agreed		# Schools: Principal & Facilitator Disagreed			# Schools: Principal & Facilitator Agreed		# Schools: Principal & Facilitator Disagreed	
	Valid N	Used	Not Used	Principal Said it was Used	Facilitator Said it was Used	Valid N	Used	Not Used	Principal Said it was Used	Facilitator Said it was Used
First come, first served	8	1	4	3	0	7	2	2	3	0
Random selection	8	0	6	1	1	7	2	2	2	1
Written application	8	1	3	4	0	7	0	5	0	2
Interviewed by Facilitator or other faculty	8	2	4	2	0	7	2	3	1	1
Efforts that reflect the range of academic abilities in the school	8	1	2	3	2	7	2	0	5	0
Efforts that reflect the ethnic makeup of the school	8	2	2	2	2	7	1	0	6	0
Preference given to students who have participated in EAST in previous years	8	2	3	2	1	7	4	0	2	1
Preference given to students based on request of parents, teachers, school counselors	8	3	3	1	1	7	4	2	1	0
Preference given to resource, special needs, or low performing students	8	1	6	0	1	7	1	5	0	1
Preference given to students for other reasons	8	0	8	0	0	7	0	7	0	0
Other	8	0	8	0	0	7	0	6	0	1
None of the above (open to all students)	8	1	6	1	0	7	0	6	0	1

While there are some ambiguities about the implications of some of the responses summarized above, they raise some important questions. Selection effects can be caused not only in the selection and placement of participants, but also by recruitment strategies. If recruitment strategies favor certain groups of students over others, or even if they simply fail to make an explicit effort to make up for the fact that certain groups of students may be less likely to apply on their own volition (i.e., to reduce self-selection), then the pool of interested students from which placements will be made may not represent the full cross section of students for whom the program is intended. For example, if it is believed that EAST is effective at increasing motivation, but recruitment relies heavily on students electing to enroll, then the less motivated students may never applyô and there is no way to know whether the program would have worked as well for those students. As shown in Table 29, some of the most commonly employed recruitment strategiesô school-wide announcements, bulletin boards, and (according to the principals), open houses for incoming gradesô would reach a very broad audience but seem unlikely to specifically encourage students who might be less inclined to apply.

However, several very targeted recruitment efforts were also commonly cited, including recommendations from the principal, guidance counselor, classroom teachers, or, most commonly, from the facilitator. Depending on what criteria these personnel used to make recommendations, however, it is unclear whether such targeted recruitment would have resulted in a more diverse group of EAST students, as advocated by the model, or a less diverse group. Disagreements between principals and facilitators about which recruitment methods were used at their schools (reported in Table 31) raise further questions about the schoolsø efforts to obtain diverse classes of EAST students.

Clearly, selection and placement procedures can also have a direct effect on the diversity of participating students. Among those strategies listed on the survey (while not exhaustive of all selection methods that could help achieve diversity), the response options that can most unambiguously be interpreted as accomplishing this include selection efforts designed to õreflect the range of academic abilities in the schoolö and to õreflect the ethnic makeup of the school.ö Although such efforts were apparently used less consistently during these schoolsø first year of operation, all responding principals indicated doing so in school year 2005-2006 (see Table 30). However, the number of facilitators who reported that such efforts were made actually decreased in the second year, and (as shown in Table 32), in 2006, only two facilitators agreed with their principals that efforts were made to reflect the range of academic abilities, and only one agreed that efforts were made to reflect the schoolø ethnic makeup. For the schools where facilitators reported different policies than the principalsô which was true in a majority of schools in the case of the two key selection strategies mentioned aboveô these discrepancies raise the obvious question of which scenarios more accurately reflect the schoolsøactual practices.

In an effort to obtain a better understanding of the reality of EAST recruitment and selection practices, questions were added to the 2006 surveys that asked the principals and facilitators about who had the primary responsibilities for these activities (the assumption being that the responsible parties would have a more accurate perspective on actual practices). The tables below summarize their responses, including the extent to which they agreed with each other about responsibilities for these activities.

Table 33Spring 2006 SurveysNumber of schools with Agreement Between Principals and Facilitators
About Recruitment Responsibilities

		Principal &	Princ Facilitator	ipal & Disagreed	
Staff Responsible	Valid N	Facilitator Both Said Staff Responsible	Principal Said Staff Responsible	Facilitator Said Staff Responsible	No One Said Staff Responsible
Principal	7	2	5	0	0
Other school administrative staff	7	0	0	0	7
EAST Facilitator	7	4	0	3	0
Counselor(s)	7	6	1	0	0
Teachers	7	0	5	0	2

Table 34

Spring 2006 Surveys

Number of Schools with Agreement Between Principals and Facilitators About Selection/Placement Responsibilities

Staff Decrease it la	Valid	Principal & Facilitator Both Said Staff	Prince Facilitator Principal Said Staff	No One Said Staff	
Stall Kesponsible	IN	Responsible	Responsible	Responsible	Kesponsible
Principal	1	1	4	0	2
Other school administrative staff	7	0	0	0	7
EAST Facilitator	7	4	1	0	2
Counselor(s)	7	7	0	0	0
Teachers	7	0	2	0	5
Placement is performed randomly by computer program	7	0	0	0	7

Table 35Spring 2006 SurveysAgreement Between Principals and Facilitators AboutFinal Decision About Placement

Staff Making Final Decision	Principal Responses	Facilitator Responses	Cited by Both Facilitator & Principal
Principal	0	1	0
Other school administrative staff	0	0	0
EAST Facilitator	7	4	4
Counselor(s)	0	3	0
Teachers	0	0	0
Placement is performed randomly by computer program	0	0	0
Total	7	8	4

As shown in Tables 33 and 34, the principals and facilitators at all seven responding schools concurred that the facilitator played a role in both recruitment and selection. However, while all seven principals also asserted that they also played a role in recruitment, only two of their facilitators concurred with this. Among the five principals who asserted that they shared responsibility for selection and placement, only one of their facilitators was aware of this. Analogously, among the five facilitators who said they played a role in recruitment and placement, only one of their principals agreed with them.

Finally, both groups were asked who made the final decision about placing a student in an EAST class in the event that there was disagreement over which students should be admitted. As shown in Table 35, all seven responding principals said that the final decision fell to the facilitator, but only four of these facilitators were aware of that.

Given that the principals themselves stated that the facilitators held greater responsibility for recruitment and selection, the fact that the facilitators did not concur about the use of selection procedures designed to reflect the ethnic makeup and range of academic abilities in the school probably means that these strategies were most likely used inconsistently, at best. This conclusion seemed to be confirmed for at least some schools by comments made by several facilitators during the spring 2006 focus group. These facilitators said that they planned to be more involved with recruitment in the following school year, now that they had a better idea of what kinds of students they wanted in their EAST classes. Several facilitators felt strongly that since participation in EAST is not a graduation requirement, it should be treated as a õprivilege,ö meriting attention to who is permitted to enroll.

Even in schools where facilitators agreed philosophically with the need to obtain a representative group of students, with so much confusion about the schoolsø policies and responsibilities for carrying them out, it is hard to imagine that they would have been very effective at creating a purposefully defined cross section of students.

The most straightforward way to determine their success at doing this would be to compare demographics of EAST students to those of other students in the school. Details of these comparative demographic analyses are shown in Appendix XX and discussed below.

These comparisons show that the EAST classes were, on average, statistically representative⁷⁴ of the remainder of the students in these eight schools in terms of the proportions of limited English proficient (LEP) students. However, the characteristics of the EAST students were different from those of their non-participating peers in many other respects, including race,⁷⁵ gender,⁷⁶ grade level,⁷⁷ Title I eligibility,⁷⁸ gifted status⁷⁹ and special education status.⁸⁰ Specifically, the EAST classes had a higher proportion of Title I eligible students overall, which

⁷⁴ Statistical comparisons based on chi-square analyses and independent samples *t*-tests had a p value greater than .05.

⁷⁵ Pearson Chi-square = 23.463, df = 4, p = .003

⁷⁶ Yates Chi-square = 3.849, df = 1, p = .050

⁷⁷ Pearson Chi-square = 131.432, df = 8, p = .000

⁷⁸ Yates Chi-square = 22.715, df = 1, p = .000

⁷⁹ Yates Chi-square =18.005, df = 1, p = .000

⁸⁰ Yates Chi-square = 21.947, df = 1, p = .000

seems to indicate that the programs have been making a conscientious effort to increase access and equity for this traditionally underserved group.

However, compared with the remainder of the students in the EAST schools, the EAST classes had more white students and fewer black students, indicating an area where the programs could do a better job of reaching out to a traditionally underserved group. EAST classes also had more students overall who had been categorized as gifted, which does appear inconsistent with the model, since the program is designed to develop the very kinds of social, problem solving, and higher order thinking skills that would be expected to improve the performance of those very students who are struggling in more traditional academic programs. The EAST programs also had fewer girls, and fewer special education students, as well as smaller proportions of middle school students who were enrolled compared with the proportions of high school students enrolled. The discrepancy in the numbers of special education students might or might not be inconsistent with the intent of the model, depending on the nature of studentsø disabilities and how these were distributed in each group. (Special education status was not examined at this level of detail.)

The above analyses show that EAST students were not representative of their peers on several potentially important demographic characteristics. However, there are many other characteristics that may also be importantô perhaps even more soô to a studentø success in EAST for which no school records exist, and which were beyond the scope of this study to assess. To the extent that EAST principals and facilitators shared the vision about the importance of such personality characteristics to the EAST program, and to the extent that they reported using these characteristics as screening criteria in ways that were consistent with the EAST model, one can feel more confident that any observed program effects can be generalized to the intended population.

Surveys were administered to all principals and facilitators in the eight EAST study schools in spring 2006 in order to determine the EAST schoolsøphilosophies about these issues. These surveys asked both groups about their schoolsø screening policies in regard to various student characteristics, including whether they made attempts to obtain certain students for EAST, to dissuade certain students from enrolling, and/or to achieve a diversity of students. In cases where they did attempt to obtain students with certain characteristics, the survey explored whether the schools tried to obtain students possessing a great deal of each characteristic, or only a small or moderate amount.

In addition, it was found during the implementation study that some schools build their EAST programs around students who represent particular demographic groups, for example, making the program available to particular grades, or targeting it to students with academic or disciplinary problems who might benefit from the programøs motivating qualities. The eight study schools were therefore also queried as to whether their programs were targeted to any particular demographic groups.

Finally, central EAST staff were also asked how *important* they felt certain characteristics were to a studentøs chances of success in EAST. This feedback was used to help prioritize the potential consequence of any discrepancies that may be found between the model

and the recruitment practices implemented in the schools. Survey results are shown in Tables 36a through 36c below.

Table 36aSpring 2006 Recruitment and Screening SurveysPotentially Helpful Student Characteristics

CHARACTERISTIC	Respondent	Level of Importance (NI=Not important; H=Helpful; VI=Very Important; Nec.=Necessary)				Screening (In=Screened in; Div.=Screened for diversity; NC=Not considered)				For characteristics that should be screened in, how much of this characteristic should a student possess?				
	Group	N	NI	Н	VI	Nec.	N	In	Div.	NC	N	Little	erate	A lot
CHARACTERISTICS THAT SHOULD BE SCREENED IN ACCORDING TO THE MODEL														
	EAST Staff	1	ç				1				1			
Previous experience with EAST	Principals						8	3	1	4	2	0	1	1
	Facilitators						7	2	2	3	2	0	1	1
	EAST Staff	1		ç			1				1	\checkmark		
Comfort with technology	Principals						8	0	4	4	0			
	Facilitators						8	1	3	4	1	1	0	0
	EAST Staff	1		Ç			1	\checkmark			1			
Self-discipline	Principals						8	0	3	5	0			
	Facilitators						8	1	2	5	1	1	0	0
	EAST Staff	1		ç			1	\checkmark			1			
Problem solving skills	Principals						8	0	4	4	0			
	Facilitators						8	1	3	4	1	1	0	0
	EAST Staff	1		Ç			1				1			
Verbal communication skills	Principals						8	0	4	4	0			
	Facilitators						8	1	3	4	1	1	0	0
	EAST Staff	1		ç			1				1			
Planning and organizational skills	Principals						8	0	3	5	0			
	Facilitators						8	1	2	5	1	0	1	0
	EAST Staff	1		ç			1				1			
Ability to focus on long-term goals	Principals						8	0	2	6	0			
	Facilitators						8	1	1	6	1	0	1	0
Willingness to take on	EAST Staff	1		ç			1				1		\checkmark	
responsibility	Principals						8	0	3	5	0			
responsibility	Facilitators						8	2	2	4	2	0	1	1

Table 36a (continued)Spring 2006 Recruitment and Screening SurveysPotentially Helpful Student Characteristics

CHARACTERISTIC		Level of Importance (NI=Not important; H=Helpful; VI=Very Important; Nec.=Necessary)					Screening (In=Screened in; Div.=Screened for diversity; NC=Not considered)				For characteristics that should be screened in, how much of this characteristic should a student possess?			
	Respondent Group	Total N	NI	Н	VI	Nec.	Total N	In	Div.	NC	Total N	Little	Mod- erate	A lot
CHARA	CTERISTICS T	HAT SHO	OULD BE	SCREE	NED IN	ACCORI	DING TO	THE M	ODEL (C	ONTINU	JED)			
	EAST Staff	1		ç			1				1		\checkmark	
Willingness to seek opportunity	Principals						8	0	3	5	0			
	Facilitators						7	1	3	3	1	1	0	0
	EAST Staff	1		ç			1	\checkmark			1	\checkmark		
Teaching/peer mentoring skills	Principals						8	0	3	5	0			
	Facilitators						7	1	2	4	1	1	0	0
Ability to collaborate and	EAST Staff	1			ç		1	ç			1	ç		
participate in teamwork	Principals						8	1	3	4	1	0	1	0
	Facilitators						8	1	2	5	1	0	1	0
Ability to work in an unstructured	EAST Staff	1			ç		1				1	\checkmark		
environment	Principals						8	0	4	4	0			
	Facilitators						8	1	3	4	1	0	1	0
	EAST Staff	1			ç		1				1	\checkmark		
Comfort in a hands-on setting	Principals						8	0	5	3	0			
	Facilitators						8	1	3	4	1	1	0	0
[For students with previous EAST	EAST Staff	1			ç		1				1		\checkmark	
experience:] Demonstrated prior	Principals						8	3	2	3	2	0	1	1
accomplishment in EAST	Facilitators						7	4	2	1	4	0	2	2

Table 36a (continued)Spring 2006 Recruitment and Screening SurveysPotentially Helpful Student Characteristics

CHARACTERISTIC		Level of Importance (NI=Not important; H=Helpful; VI=Very Important; Nec.=Necessary)					Screening (In=Screened in; Div.=Screened for diversity; NC=Not considered)				For characteristics that should be screened in, how much of this characteristic should a student possess?				
	Respondent Group	Total N	NI	Н	VI	Nec.	Total N	In	Div.	NC	Total N	Little	Mod- erate	A lot	
CHARACTERISTICS THAT SHOULD BE SCREENED FOR DIVERSITY ACCORDING TO THE MODEL															
	EAST Staff	1	ç				1		\checkmark		0				
Academic giftedness ó math	Principals						8	0	1	7	0				
	Facilitators						8	0	2	6	0				
	EAST Staff	1	ç				1		\checkmark		0				
Academic giftedness ó literacy	Principals						8	0	1	7	0				
	Facilitators						8	0	2	6	0				
	EAST Staff	1	ç				1		\checkmark		0				
Academic giftedness ó science	Principals						8	0	1	7	0				
	Facilitators						8	0	2	6	0				
	EAST Staff	1	Ç				1		\checkmark		0				
Written communication skills	Principals						8	0	4	4	0				
	Facilitators						8	1	3	4	1	1	0	0	
	EAST Staff	1		ç			1		ç		0				
Desire to change the community	Principals						8	0	1	7	0				
	Facilitators						8	1	1	6	1	1	0	0	
	EAST Staff	1		ç			1		\checkmark		0				
Ability to work independently	Principals						8	1	3	4	1	0	0	1	
	Facilitators						8	2	1	5	2	0	1	1	
	EAST Staff	1		ç			1				0				
Leadership ability	Principals						8	0	2	6	0				
	Facilitators						8	1	2	5	1	1	0	0	
	EAST Staff	1		ç			1				0				
Willingness to work hard	Principals						8	1	3	4	1	0	1	0	
	Facilitators						8	2	2	4	2	1	1	0	

Table 36a (continued)Spring 2006 Recruitment and Screening SurveysPotentially Helpful Student Characteristics

CHARACTERISTIC	Beenendont	Level of Importance (NI=Not important; H=Helpful; VI=Very Important; Nec.=Necessary)					Screening (In=Screened in; Div.=Screened for diversity; NC=Not considered)				For characteristics that should be screened in, how much of this characteristic should a student possess?			
	Group	Total N	NI	н	VI	Nec.	Total N	In	Div.	NC	Total N	Little	Mod- erate	A lot
CHARACTERISTICS THAT SHOULD BE SCREENED FOR DIVERSITY ACCORDING TO THE MODEL (CONTINUED)														
	EAST Staff	1		ç			1				0			
Ambition	Principals						8	0	2	6	0			
	Facilitators						8	1	2	5	1	1	0	0
Salf accurance	EAST Staff	1		Ç			1		√ 2		0			
Sen-assurance	Principals						8	01	2	6 7	0			
	Facilitators						0	1	0	/	1	1	0	0
Literacy skills (as specified in the Arkansas English Language Arts	EAST Starr	1		Ç				0	N 4	4	0			
Curriculum Frameworks)	Fincipals						0 0	1	4	4	1			
	Facilitators						0	1	3	4	1	1	0	0
Mathematics skills (as specified in the Arkansas Mathematics	EAST Starr	1		Ç			1 0	0	N 2	5	0			
Curriculum Frameworks)							<u> </u>	1	3 1	5	1			
, 	Facilitators						8	1		6	1	1	0	0
Science knowledge and skills (as specified in the Arkansas Science	EAST Staff	1		Ç			1	<u> </u>	ν 2	5	0			
Curriculum Frameworks)	Frincipals						<u>ð</u>	1	3 1	5	1			
, 	Facilitators						8	1	1	0	1	1	0	0
Social studies knowledge and skills (as specified in the Arkansas Social Studies Curriculum Frameworks)	EAST Staff	1		Ç			1	0	ν 2	5	0			
	Principals						8	0	3	ے ب	0			
	Facilitators						8	1		6	1	1	0	0
English language proficiency	EAST Staff	1			Ç		1	0	N 4	4	0			
English language proficiency	Frincipals						8 0	0	4	4 5				
	racilitators						ð	U	3	3	U	U	U	0

Table 36bSpring 2006 Recruitment and Screening SurveysPotentially Hindering Student Characteristics

CHARACTERISTIC		Level of Importance (NI=Not important; H=Helpful to Avoid; VI=Very Important to Avoid; Nec.=Necessary to Avoid)					(In=S Div.=)	creened Screened co	Screening in; Out= l for dive onsidered	g Screene ersity; N(1)	For characteristics that should be screened in, how much of this characteristic should a student possess?				
	Respondent Group	Total N	NI	Н	VI	Nec.	Total N	In	Out	Div.	NC	Total N	Little	Mod- erate	A lot
	CHARAC	TERIST	ICS THA	T SHOU	LD BE S	CREENE	D OUT A	ACCORI	DING TO	THE MO	DDEL				
Doesnøt want to be there	EAST Staff Principals Facilitators	1 			ç 		1 8 7	2 0	$\frac{\sqrt{0}}{1}$	0	6 5	0 2 0	 1 	 0 	 1
Physical disability that can	EAST Staff	1				ç	1	0		4	4	0			
assistive technologies	Facilitators						8 7	0	0	4 3	4	0			

Table 36b (continued)Spring 2006 Recruitment and Screening SurveysPotentially Hindering Student Characteristics

CHARACTERISTIC		(NI= Avoid	Level of ImportanceScreening(NI=Not important; H=Helpful to Avoid; VI=Very Important to Avoid; Nec.=Necessary to Avoid)(In=Screened in; Out=Screened of Div.=Screened for diversity; NC= 						d out; C=Not	<i>t;</i> should be screened in, how much of this characteristic should a student possess?					
	Respondent Group	Total N	NI	Н	VI	Nec.	Total N	In	Out	Div.	NC	Total N	Little	Mod- erate	A lot
CHARACTERISTICS THAT SHOULD BE SCREENED FOR DIVERSITY ACCORDING TO THE MODEL															
Disciplinary problems	EAST Staff Principals Facilitators	1 	ç 				1 8 7	0	2 3	$\frac{\sqrt{3}}{2}$	32	0 0 0			
Academically struggling	EAST Staff Principals Facilitators	 	ç 	 			1 8 7	0	2 1	$\frac{\sqrt{1}}{3}$	5 3	0 0 0			
Doesnøt take school seriously	EAST Staff Principals Facilitators	1 	Ç 				1 8 7	2 0	0 √	$\frac{\sqrt{0}}{2}$	6 4	0 2 0	 1 	 0 	 1
Physical disability that can be overcome with assistive technologies	EAST Staff Principals Facilitators		ç 				1 8 7	1	0	$\frac{\sqrt{4}}{4}$	3	0 1 0	 0 	 1 	 0
Emotional problems	EAST Staff Principals Facilitators	1 		ç 			1 8 7	0	1	$\frac{\sqrt{2}}{2}$	5	0 0 1	 0	 1	 0
Socialization problems	EAST Staff Principals Facilitators			ç 			1 8 7	0 1	1	$\frac{\sqrt{2}}{3}$	5 3	0 0 1	 0	 1	 0
Short attention span	EAST Staff Principals Facilitators			ç 			1 8 7	0	1	$\frac{\sqrt{2}}{2}$	5 5	0 0 0			
õGrade seekerö	EAST Staff Principals Facilitators	1 		ç 			1 8 7	0	0	$\frac{\sqrt{3}}{2}$	5 5	0 0 0			
Lack of motivation	EAST Staff Principals Facilitators	1 		ç 			1 8 7	1 0	0	$\sqrt{\frac{3}{3}}$	4	0 0 0			

			For each chara	cteristic, indicate if a	school should:
CHARACTERISTIC	Respondent Group	Total N	Recruit or select students from particular subgroups	Recruit or select students for diversity on this characteristic	Not consider this characteristic
	EAST Staff	1			
Gender	Principals	8	0	6	2
	Facilitators	8	0	5	3
	EAST Staff	1		\checkmark	
Age	Principals	8	1	5	2
	Facilitators	8	2	3	3
	EAST Staff	1		\checkmark	
Grade level	Principals	8	3	4	1
	Facilitators	8	4	3	1
	EAST Staff	1		\checkmark	
Race/ethnicity	Principals	8	0	6	2
	Facilitators	8	0	6	2
English languaga	EAST Staff	1		\checkmark	
proficiency (LEP Status)	Principals	8	0	6	2
proneiency (LEI Status)	Facilitators	8	0	4	4
Disabled/special	EAST Staff	1		\checkmark	
education	Principals	8	0	6	2
education	Facilitators	8	0	6	2
	EAST Staff	1		\checkmark	
Socioeconomic status	Principals	8	0	б	2
	Facilitators	8	0	6	2

Table 36cSpring 2006 Recruitment and Screening SurveysDemographic Characteristics

As shown at the top of Table 36a above, among the student characteristics listed, central EAST staff felt that there were 14 characteristics that schools should look for, at least to some extent, in potential EAST students (characteristics that should be õscreened inö). Among these, there were four characteristics that the staff considered to be *very important* for a candidate to possess at least to some extent, including the ability to collaborate and participate in teamwork, the ability to work in an unstructured environment, comfort in a hands-on setting, and (for students with previous EAST experience), demonstrated prior accomplishment in EAST. However, all or almost all of both principals and facilitators indicated either that these characteristics were *not considered* as screening criteria, or that their screening policies were designed to achieve a range of these characteristics among students. While EAST staff considered it important that students possess these characteristics only to a small extent, the implication is that schools seeking diversity in these areas might have been accepting students who did not possess these characteristics at all, while those schools that did not even consider these as screening criteria certainly could not ensure their presence.

In addition, the lower part of Table 36a lists another 15 student characteristics that the EAST staff felt should be included in the definition of diversity when seeking to obtain EAST classes that were representative of the school as a whole. None of these characteristics were cited by a majority of principals or facilitators as among those on which their schools sought to

establish diversity, and a substantial number of them were not considered in the screening process in most schools.

Similarly, there are also a number of characteristics that EAST staff believed could hinder a student¢s potential for success (Table 36b) but that EAST is believed to help students overcome, especially if the class is heterogeneous. These characteristics are listed in the lower part of Table 36b. As for the favorable characteristics, none of these student attributes were cited by a majority of principals or facilitators as among those on which their schools sought to establish diversity. In many cases, the majority of schools indicated that these variables were not considered during screening, and it is likely that many respondents intended by this that they do not count such characteristics against a candidate when considering placements. However, as discussed previously in the discussion of recruitment, casting a wide net and adopting policies that are õblindö to such attributes are not sufficient to insure diversityô much less representativenessô in the classroom.

Most schools *did* report that they seek diversity on traditional demographic variables such as gender, race, and socioeconomic status (see Table 36c); but seeking diversity is not necessarily the same as seeking representativeness. While the EAST classes did include highly diverse populations, they were often not representative of the rest of their schools, as discussed above. In addition, since many of the characteristics on which schools did not explicitly screen for diversityô such as desire to change the community, ability to work independently, leadership ability and the likeô have not been measured and are not formally documented, it is not possible to confirm empirically whether or not EAST classes were representative of their schools in these respects, although classroom observations and anecdotal evidence does seem to indicate diversity in these regards. In some classes, observers noted extensive diversity in terms of studentsø apparent motivation levels, leadership skills, and work ethics. In other observed classes, however, students appeared much more homogeneous on characteristics such as motivation. These observations were inconclusive, however. They were not necessarily representative of all EAST classes in the study schools, and even to the extent that diversity was observed, there was no way for observers to know whether it was representative of the diversity in the school as a whole. In the end, the fact that explicit efforts were rarely made to ensure diversity and representativeness raises the possibility that such efforts may not have been the norm.

As mentioned at the beginning of this section, this situation does not necessarily indicate significant problems with program implementation, particularly with schools in the early stages of the program. However, it does have potential implications with regard to study findings about program impact; these implications are discussed further under Conclusions.

Student Outcomes

Among the broad research questions that were discussed in the introduction to this report, the specific outcomes that were explored in the outcomes study, as described in greater detail in the Procedures section above, were academic performance, motivation for school, problem solving skills, self-directed learning style, and motivation to pursue further learning. These outcomes were assessed on the Iowa Tests of Basic Skills and Iowa Tests of Educational Development, Inventory of School Motivation, Social Problem Solving Inventory for Adolescents, and the student survey, respectively. The impact of the EAST program on each of

these outcomes, after controlling for various student characteristics and environmental variables, was measured using hierarchical linear modeling, also described in detail under Procedures. All of these impacts were further explored through qualitative methods including individual and focus group interviews, classroom observations, and surveys, all of which have been discussed above. Findings from these analyses are discussed in the following sections.

Implications of Student Mobility and Instrument Response Rates

An important consideration for all outcome measures is whether any changes in the population (due to attrition, non-response, etc.) that occur after the samples are selected might make the EAST and control groups less similar, or result in misleading findings. The first question is whether program attrition (due to students dropping the course, changing schools, or dropping out of school) might have biased the study conclusions.⁸¹ For example, it is possible that if the less academically successful students in the EAST classes are more likely than others to drop the class or drop out of school, it could make the program appear more successful than it really is. Indeed, there was consensus among EAST students participating in the 2006 focus group that at times there had been students who were õless likely to try new thingsö and õdidn¢t like having to think for themselvesö who were more likely to drop the class.

Since most of the outcome measures were obtained from surveys and inventories that could not be administered to students who left the school, it was not possible to assess these possibilities directly by examining outcomes for the students who left. However, to the extent that the degree of mobility was meaningful, it is possible to establish whether it appears likely to have resulted in misleading conclusions by comparing other characteristics of students who left the program or the school (õleaversö) to those who remained (õstayersö). Among the 509 students on the original rosters of students in EAST classes during the 2005-2006 school year in the participating study schools, 19 (4%) moved to another school at some point during the school year (õleaversö), 7 (1%) dropped out of school (or disappeared from the stateøs records), and the remaining 95% remained in the school through the end of the school year (õstayersö).⁸² Mobility rates are not applicable to control students, since the control group was selected after the end of the school year from among all students who had remained in the school (and who at least had fall test data available).

While these attrition rates were modest, it is nevertheless potentially important to compare the characteristics of those who left the school to those who remained. Detailed results of these comparisons for EAST students are provided in Appendix XXI. The analyses revealed that, among EAST students in the study schools, there were no statistically significant differences in mobility rates of students of different genders, races, grade levels, nor based on Title I, gifted or special education status, nor were the fall 2005 ITBS reading and math scores

⁸¹ Since the Cohort 2 control group was selected after the end of the school year, it was possible to create a comparison group *post facto* that was comparable to the original target population and from which there was no attrition.

⁸² Note that some of those who remained in the school may have dropped the EAST class; however, the ADE did not have records available of mid-year changes in course enrollments. This would have resulted in our underestimating both the mobility rates and the response rates.

significantly different for leavers, stayers or drop-outs.⁸³ However, there was a significant difference in the mobility rates of LEP students compared with English proficient students: one of the seven ELL students in the EAST classes (14%) dropped out of school, compared with only 1% of the non-ELL students. While these differences were statistically significant,⁸⁴ however, the number of ELL students in EAST was so low (representing only 1% of the population of the EAST classes in the study schools) that the discrepancy does not present a problem for the validity of the study.

While mobility does not appear to have posed a meaningful threat to the validity of the outcomes study, additional attrition from the analyses occurred as a result of the fact that not all students completed the reading and math tests and/or student inventories. This raises the question of whether the response rates obtained on the measurement instruments were high enough to be reasonably confident that the conclusions would have been unlikely to have changed had the results included feedback from the non-respondents. (This could happen, for example, if there were a substantial number of non-respondents, and those students who made the smallest pre/post gains were least likely to respond, in which case the observed mean gains would appear greater than they really were. Or, if students entering the program with the lowest skills were less likely to return a pretest but made substantial gains during the year, it would make the observed mean gains appear smaller than they really were.) In addition to course completers who did not return tests or inventories, some students may have dropped the EAST course during the year. While the number dropping the course is not known because course enrollment records were not available at that level of detail, these students would not have completed the spring inventories and would therefore be reflected in the response rate analyses.

In cases where response rates are less than optimal, we can project whether they appear likely to result in misleading conclusions by comparing respondents and non-respondents in terms of other characteristics such as demographics and prior academic performance (where available). These comparisons were conducted separately for EAST and control students in order to determine whether there was any apparent response bias in either group.

Response rates for EAST students were calculated on the three outcome instruments⁸⁵ based on all students who remained in the EAST class for the full year and therefore had an opportunity to respond to both fall and spring instruments. Specifically, this included all students on the original EAST rosters, less those who pre-registered but never enrolled, dropped the course, moved to a different school, dropped out of school, or disappeared from the system at some time during the school year. Since there was no mobility among the Year 3 control students as explained above, response rates for this group were calculated for all selected control students. For instruments administered on a pre-post basis (including the ITBS/ITED and the motivation and problem solving inventories), only those students who responded in both fall and spring were counted as respondents. Response rates are summarized in Table 37 below.

⁸³ While the mean scores for drop-outs were lower than for the other groups, there were so few drop-outs that the difference was not statistically significant.

⁸⁴ Pearson Chi-square = 8.930, df = 2, p = .012

⁸⁵ Reading and math scores were derived from the same instrument, as were motivation and problem solving scores, so only three instruments were returned for the five outcomes.

Resp	onse Rates Amo	ong EAST and Col	ntrol Students
Target Group	# of "Stayers"	Instrument	# (%) Matched Responses
		ITBS/ITED	420 (87%)
EAST	483	Student Inventory	367 (76%)
			430 (89%)
		ITBS/ITED	457 (92%)
Controls	496	Student Inventory	391 (79%)
		Student Survey	424 (86%)

 Table 37

 Response Rates Among EAST and Control Students

As the above data show, response rates among both EAST and control students were good (above 75%) for all three instruments, and were especially high on the ITBS/ITED and on the student survey, even after accounting for students who may have dropped the class. However, there was a modest but statistically significant difference between the groups in response rates on the ITBS/ITED,⁸⁶ with slightly higher response rates for control students than for EAST students (92% vs. 87%).⁸⁷

Because of the difference in ITBS/ITED response rates between EAST and control students, demographics were compared between respondents and non-respondents on each instrument to determine whether the missing responses and/or the differences in response rates between the two groups may have biased the sample. Detailed results of these comparisons are presented in Appendices XXI and XXII.

For the ITBS/ITED, there were no significant differences between respondents and nonrespondents in the EAST classes in terms of their racial distributions or their gifted, special education, or ELL designations. However, there was a larger proportion of boys (70% vs. 54%),⁸⁸ a larger proportion of high school students (68% vs. 44%),⁸⁹ and a *smaller* proportion of Title I students (71% vs. 86%)⁹⁰ among non-respondents than among respondents. For control students, there were no significant differences between respondents and non-respondents in terms of the proportions of boys and girls, or of their gifted or special education designations or Title I status. However, as for EAST students, there was a larger proportion of high school students among non-respondents than among respondents (56% vs. 46%).⁹¹ For control students, there was also a larger proportion of Hispanic students (15% vs. 3%)⁹² and a much larger proportion of ELL students (15% vs. less than 1%)⁹³ among non-respondents.

The final samples for whom data were available and that were used for the analyses of the ITBS/ITED outcomesô i.e., those students who returned both pre and posttestsô were

⁸⁶ Yates Chi-square = 6.493, df = 1, p = .011

⁸⁷ This difference resulted in part from the fact that availability of a pretest score on the ITBS/ITED was one of the criteria used in the selection of control students. This was necessary in order to insure that the control group was as similar as possible to the EAST students, including in their academic performance.

⁸⁸ Yates Chi-square = 5.080, df = 1, p = .024

⁸⁹ Pearson Chi-square [grade level X response group] = 28.869, df = 6, p = .000

⁹⁰ Yates Chi-square = 7.574, df = 1, p = .006

⁹¹ Pearson Chi-Square = 15.537, df = 6, p = .016

⁹² Pearson Chi-Square [race X response group] = 16.370, df = 4, p = .003. Hispanic was the only race category where there was a substantial difference.

⁹³ Yates Chi-square = 49.000, df = 1, p = .000
compared to determine whether they remained comparable after the effects of attrition and nonresponse. These resulting samples were statistically indistinguishable in terms of most descriptive variables, including mean reading and math pretest scores and distributions by gender, grade level, and gifted, special education and ELL designations. However, there were slightly more white students $(73\% \text{ vs. } 61\%)^{94}$ and slightly more Title I students $(90\% \text{ vs. } 86\%)^{95}$ in the control groupô reflecting differences in the original samples. Details of these comparisons are presented in Appendix XXIV.

Similar analyses were conducted for the Student Inventory respondents, who formed a mostly overlapping but slightly different population than the ITBS respondents since respondents did not always return both instruments. As shown in Appendix XXV, comparisons of the resulting EAST and control samples were very similar to those for the ITBS respondents. Among inventory respondents, EAST and control students were also statistically indistinguishable in terms of their mean reading and math pretest scores and distributions by gender, grade level, and gifted, special education, and ELL designations. However, there were slightly more white students (75% vs. 63%)⁹⁶ and slightly more Title I students (89% vs. 84%)⁹⁷ in the control group.

While some of these differences were statistically significant, they were not very substantial, and since all of these demographic factors were included as control variables in the HLM analyses, it is not expected that they would cause problems for the validity of the study.

Impact of EAST on Student Skills and Content Knowledge

As described in Section a) above, statistical models were built using hierarchical linear modeling in order to determine whether any effects were found for the EAST program after controlling for other variables that might also influence these outcomes. These analyses also provided insight into the extent to which EAST may work better for some students, or in certain environments or contexts, than for others.

Among the 16 outcome variables that were tested (including the four major areas of academic achievement, school motivation, problem solving skills, learning style, and plans for after high school, as well as several sub-skills within problem solving and motivation), the HLM analyses revealed a statistically significant relationship to participation in EAST in five domains, summarized in the following table.

⁹⁴ Pearson Chi-Square = 14.701, df = 4, p = .005

⁹⁵ Pearson Chi-Square = 3.981, df = 1, p = .046

⁹⁶ Pearson Chi-Square = 18.559, df = 4, p = .001

⁹⁷ Yates Chi-Square = 4.032, df = 1, p = .045

Domains on which Relationship to EAST 1 articipation was Found		
Instrument	Domain	Description
SPSI-A	Problem Identification	Skill at defining the characteristics of a problem
SPSI-A	Problem Evaluation	Skill at assessing the outcomes of a solution
SPSI-A	Problem Reorganization	Skill at revising strategies in response to the assessment
		of outcomes
ISM	General Mastery	Motivation for school derived from accomplishment
Student Survey	Learning Style	Self-directed learning style

 Table 38

 Domains on Which Relationship to EAST Participation Was Found

Detailed results of the HLM analyses for each of the 16 outcome areas are presented in Appendix XXVI.

While all of the tested outcome variables are relevant to the goals of the program, the preponderance of evidence for program effects in the area of problem solving skills seems consistent with one of the most central goals of EAST, and may point to a particular strength of the program. After all, problem solving, in a variety of forms, is what students *do* in EAST on a daily basis. Indeed, during the implementation study focus groups, facilitators had volunteered that EAST develops studentsø practical skills such as familiarity with real world problems and how to solve them, including the ability to find their way around community organizationsø procedures and organizational structures in working towards project goals. Students from the study schools also concurred that an inclination to õseek solutionsö is among the qualities that EAST develops.

It is worth reiterating here that the problem solving construct that was tested in this studyô and that EAST strives to buildô is not limited to the traditional, narrow definition of the concept that relates primarily to math problems or logic puzzles. Rather, the õsocial problem solvingö construct refers to a broad set of õattitudes, behaviors and skillsö for a process that õis taking place in a real-life situationö and õis part of a series of cognitive and behavioral skills related to social learningö (Frauenknecht and Black, 1995). It might be argued that, in developing their approach to problem solving, students are not merely learning a skill, but learning to learn.

In the current analyses, the best-fitting statistical model for *Problem Identification* (Appendix XXVI, model 11) suggested that participation in a high school EAST program had a positive influence on studentsø ability to identify and define problems. Higher problem identification skills on the pretest were also found (unsurprisingly) to be associated with higher skills on the posttest. More interestingly, their skills in this area were also found to be positively influenced by their initial problem evaluation and math skills, and by the extent to which they were motivated by accomplishment and by working with and for the benefit of others (*ISM* – *General Social*).⁹⁸ However, initial evaluation skills were found to have a greater impact on

⁹⁸ While there was no statistically reliable evidence of a direct impact of EAST on *General Social* motivation, there was a clear consensus among EAST students who had been interviewed over the years that social skills involving working with othersô such as becoming more outgoing, learning to speak in front of others, willingness to meet people, increased familiarity with the community, and leadership and consultation skillsô were among the benefits of participation.

problem identification skills for students in control schools, or in EAST schools with lower fidelity ratings, than in the higher fidelity programs. While it is not specifically demonstrated by the analysis results, one possible explanation for this might be that the best EAST programs were able to provide a form of support that compensated for a lack of initial problem evaluation skills.

Consistent with the relationship between program fidelity and initial problem evaluation skills discussed above, participation in an EAST high school program was also found to have a positive influence on studentsø final *Problem Evaluation* skills (Appendix XXVI, model 15). Specifically, for students who were initially unskilled at generating alternative solutions to a problem (*SPSI-A – Alternative Generation*), participation in a high school EAST program had a positive influence on their ability to assess the outcomes of an attempted solution (*Evaluation* subscale). In addition to the expected finding that higher evaluation skills on the pretest were predictive of higher skills on the posttest, the development of problem evaluation skills was also positively influenced by a studentøs initial reading skills. In addition, students who were initially more motivated by working hard (*ISM Effort*) and by accomplishment (*ISM General Mastery*) prior to starting the program also developed stronger problem evaluation skills.

The third aspect of problem solving that EAST was found to develop was Reorganization skills (Appendix XXVI, model 16). Participation in EAST was found to have a positive influence on studentsø ability to revise strategies in response to the assessment of outcomes, and length of participation (first-year Cohort 2 students vs. Cohort 1 students who were repeating the program) was also found to make a difference. Studentsø development of problem reorganization skills was also positively influenced by their initial Problem Identification and Alternative Generation skills. Those who were more motivated by working hard (ISM – Effort) and by working with and for the benefit of others (ISM - General Social) prior to starting the program also developed stronger reorganization skills. While lower socioeconomic status (eligibility for free or reduced price lunch) proved to be a statistically significant handicap for the development of reorganization skills, it did not attenuate the benefits of the EAST program in this domain. Surprisingly, higher reorganization skills on the pretest were not found to be predictive of higher skills on the posttest. This might be because *Reorganization* was so closely related to another variable or variables that were included in the model that its influence could not be separated out.

In the area of school motivation, EAST was found to help students develop motivation for accomplishment (ISM – *General Mastery*; see Appendix XXVI, model 7), and length of participation was again also found to make a difference. This type of motivation was also found to be influenced by the overall *Problem Solving Skills* that a student possessed prior to entering the program. In addition, those who were initially motivated by hard work also developed stronger motivation for accomplishment. Similarly, the extent of a student¢s initial motivation derived from praise (ISM – *Praise*) was associated with gaining stronger motivation for accomplishment, but this was less true for low-income students.

Finally, participation in the EAST program was found to increase the extent to which a student reported independence, resourcefulness, and leadership in their work style. This construct, which was measured on the spring 2006 student survey, was computed by taking the mean of five items, each measured on a 5-point scale (where 1 = strongly disagree and 5 =

strongly agree) addressing studentsø perceptions of their learning styles. The items included were as follows:

- When working on group projects, I can clearly explain my role in the group to others.
- I am able to take a leadership role in group activities.
- I know of people I can ask for help with school work other than my teachers or parents.
- If I have to, I can usually complete a school project without much help from my teachers.
- When describing a project Iøm working on, I can change the way I explain it to make it clearer if the people Iøm talking to donøt understand.

We refer to this construct as self-directed learning style⁹⁹ (see Appendix XXVI, model 9). Students in EAST schools with higher fidelity programs were found to develop a slightly (but statistically significantly) stronger self-directed learning style than those in lower fidelity programs or in control schools. In addition, students who were initially motivated by hard work and by working with and for the benefit of others were also found to develop a stronger self-directed learning style. A very slight (but statistically significant) positive correlation was also found with studentsøinitial math skills.

Although no direct effects were found indicating an impact of the EAST program on studentsø math and reading test scores, this is a notoriously difficult relationship to demonstrate. Given the myriad of other factors that influence academic achievement (which could never realistically all be controlled for in a study of this kind), and the limitations of standardized testing for measuring such skills, this should be taken as a failure to find a relationship, but certainly not as evidence that none exists. However, the consensus in the literature (as discussed in the Research Context section) of the importance for academic and career success of the types of behavioral and attitudinal characteristics that have been explored in this study lends considerable additional weight to the fact that EAST *has* been shown to support studentsø development in several of these important areas.

Conclusions

Participation in the EAST program was found to have a significant positive impact on several outcome areas that are central to the objectives of the program, with a particular emphasis on developing studentsø problem solving skills. Specifically, EAST was observed to improve participantsø skills, or at least their awareness of what is involved in defining the characteristics of a problem, assessing the outcomes of a solution that was implemented, and revising their strategies in response to the assessment of those outcomes. EAST was also found to increase the degree to which students derived motivation for school accomplishment and achieving success in their studies, and the extent to which they felt comfortable with a self-directed approach to their studies. These findings were consistent with comments that have been made by students and facilitators during focus groups throughout the three-year study. These individuals have reported that the program changes studentsø õsocial experienceö by improving

⁹⁹ This scale was constructed *a priori* by composing survey items that the EAST and ADE staff agreed were appropriate for measuring an important outcome of the program. While it was not verified through factor analysis, it does have strong face validity, and was also found to have good internal consistency (Cronbachøs = .796).

their discipline, initiative, and sense of responsibility, and it better enables them to take charge of their own educational experience. Focus group participants also expressed consensus about the value of the EAST class in helping them with their personal goals. By providing opportunities to õsee different things,ö both middle and high school students felt that EAST is helping them identify their goals in the first place. Because the class boosted their confidence and helped them recognize their potential, they set higher goals for themselves. In addition, they indicated that the classøs flexibility and the broad range of new experiences to õexplore the entire worldöô for example, in areas such as architectural design, computer programming, and web designô are opening up new possibilities and helping them move towards their goals, in addition to helping them develop technical, communication, and interpersonal skills that they felt would be valuable for career skills.

Pertinent to the question of what types of students EAST is good for, there were a number of factors that were found to reduce the likelihood of success in achieving these objectives, but findings indicated that while they would put a student at a disadvantage, for the most part they were *not* observed to reduce the magnitude of the impact of EAST participation. Specifically, among the five domains that were found to be impacted by EAST, coming from a low income family was found to put students at a disadvantage for development of their reorganization problem solving skills, while students in urban schools were disadvantaged in their alternative generation skills compared with their peers in rural schools. In addition, while students who were initially motivated by receiving external praise tended to develop greater motivation from accomplishment, this was less true for low income students who were motivated by praise. In all of these cases, however, these characteristics did not reduce the impact of the EAST program. This last finding confirms the program staff¢s and students¢ observations that motivated students are more likely to succeed in EAST. Perhaps more importantly, however, it also indicates that EAST can still be beneficial in this domain even *in spite* of low initial motivation.

The one notable exception where EAST was found to have provided less benefit for certain students was that, while participation was found to have an impact on developing studentsø problem evaluation skills, it only proved to be beneficial in this domain for students who were initially unskilled at alternative generation. This should not be taken to mean, however, that students who *are* already skilled at alternative generation cannot benefit from the program, since they could still benefit (at least) in other domains (although they may be less likely to improve their alternative generation skills due to ceiling effects).

It may be that there are other variables that were not measured in this study that could create an obstacle to success in EAST. However, among the student characteristics that were studied, the fact that there was only one that was specifically found to reduce the programøs benefit in any wayô and in that case, it was reduced in only one domainô further underscores the importance of involving a diversity of students in the program. However, our efforts to determine the recruitment and selection procedures that schools were using and, by extension, to obtain an understanding of the characteristics of the population of EAST participants, encountered unexpected difficulties in this study. The fact that principals and facilitators so often provided inconsistent information about their schoolsø policies raise questions about what actually took place, how consistently it took place, and how effective any efforts the schools may

have been making to encourage diversity might have been. This could have implications for the studyøs conclusions, as discussed further under Limitations, below. While the eight fledgling study schools are likely not, on the whole, representative of all EAST schools, it might also mean that EAST might not have been reaching as many lower-motivated students as it couldô at least in the early years at these schoolsô because such students are probably less likely to seek the program out. Given that EAST has been shown in this study to have a positive effect on studentsø motivation for school, it is appropriate that project leadership continue to work with new and established schools to emphasize the importance of actively seeking out diversity.

Although it is not necessarily a goal of the program to involve students for more than one year (with the exception of a handful of experienced students who should be used to õseedö the new cohort each year), the fact that length of involvement (cohort) was found to have a differential impact in at least two outcome areas (motivation from accomplishment and development of reorganization problem solving skills) raises the possibility that a longer stay in EAST might be beneficial in other ways as well. Indeed, the fact that length of involvement was found to be beneficial at all is particularly significant in light of the fact that two of the eight schools experienced setbacks during the study resulting from reconfiguration of the schools and/or changes in staff, without which they might have advanced even further than they did.

Similarly, the EAST director has observed that it normally takes from three to five years for a program to become fully established and owork out the kinks.o This observation is consistent with the literature about the length of time needed to achieve systemic change. This study explicitly involved only new EAST programs, so as to minimize the number of unobserved historical factors that might have influenced the outcomes. Given that these programs were really still just starting up, and that several of them had not yet reached a high level of program fidelity, the fact that measurable program impacts were still observed is all the more impressive.

This raises the possibility, however, that the fidelity measure might have been biased by the decision to use ADE and EAST, Inc. staff as observers, as one might suspect that vested program staff might be less objective than an outside observer. In truth, this decision represented an unavoidable trade-off. We believe that the integrity of these staff (who have never hesitated to raise concerns about program implementation with school personnel), combined with a rigorous training process which included raising their awareness that anything other than accurate and objective observation ratings would have an unpredictable effect on the ability to measure program impacts, were able to achieve objective ratings. Nevertheless, their involvement introduces the appearance of the possibility of bias. This possibility was a necessary cost for meeting the financial and pragmatic necessity of having a sufficient number of observers to conduct a large enough number of observations. It was also a trade-off with the need to use observers with a strong understanding of the program in order to achieve a sufficiently rigorous rating process. In this sense, the trade-off between the appearance of reduced objectivity and the very real need to ensure accuracy was deemed worthwhile.

While the fidelity measure appears to have been successful at separating higher and lower functioning programs, at least in terms of overall fidelity, it is unfortunate that it did not achieve better discriminant validity. With the ability to control for specific program strengths and weaknesses by assessing fidelity more precisely within specific program domains, it is quite

possible that even stronger findings, or findings in additional outcome areas, might have been observed.

Limitations

While every effort was made during the study to maintain the integrity of the evaluation, it is inevitable when conducting research in public school settings that conditions will fall short of the ideal. Following are some of the most significant limitations of this study, with speculation about their potential implications.

Probably the most important limitations have to do with the implications of the student Since the process of selecting control students was very successful in selection process. obtaining a group that was very similar to the target EAST students, and these groups remained comparable even after school attrition and instrument non-response were taken into account, our conclusions about differential outcomes for EAST students compared with control students should not have been biased by systematic differences between the two groupsô at least not in terms of the kinds of measurable demographic characteristics that were used in the selection process (and also controlled for statistically). However, the observed demographic differences between EAST students and other students in their schools, along with principalsø and facilitatorsøresponses to the recruitment surveys, underscore that the schools did not consistently create EAST classes that were representative of the student body. While we do not presume to judge the appropriateness of such recruitment policies from a pedagogical perspective, the implication for the research is that there is a distinct possibility that EAST students may differ from the control students in important ways. This can happen even in spite of the purposive matching procedure that was used to obtain a comparable group of control students, since the groups could differ on characteristics that were not measured.

Such differences could have implications for the generalizability of the study, and might even compromise the validity of the findings, depending on the specific characteristics (if any) on which students differed. For example, study results would not be expected to generalize to students who would be screened out from a õmodelö program (e.g., students who did not want to be there), or who do not possess at least some degree of the characteristics that should be screened in (such as self-discipline, willingness to take on responsibility, basic problem solving, and verbal communication skills). However, most schoolsørecruitment and selection processes did not actively seek diversity on a number of helpful characteristics that were not cited by the model as characteristics that should be screened in, nor on several hindering characteristics that were not cited as among those that should be screened out. Since most schools said they did not consider these characteristics in recruitment and screening, one can not rule out the possibility that students who possess the positive characteristics mentioned and/or those who do not possess the hindering characteristics are more likely to seek out the program. This raises the possibility that the program impacts for students who do not resemble the latter group when they enter the program might be smaller than the impacts observed in this study. Indeed, this would be consistent with comments made by students as well as facilitators, both of whom described a number of prior characteristics that they felt make a student more likely to succeed in EAST. These included, for example, a desire to change the community, an ability to collaborate and participate in teamwork, an ability to work independently, comfort in non-traditional (hands-on)

educational settings, a willingness to work hard, and ambition. While these characteristics may increase a studentøs likelihood of success, those who do not already possess these characteristics might also achieve success in EASTô and in fact anecdotes abound about cases where this has happened.

A related issue is that while schools were randomly assigned as EAST or control schools from among those applying to the program in spring 2004, and EAST and control students were found to be quite similar, this pool of applicants may not have been representative of all EAST schools, much less of all Arkansas schools. While the 16 study schools did not appear to be particularly unusual in any obvious ways, since the program is voluntary, it is likely that there is a self-selection effect among schools, just as among students. For this reason, one can not safely conclude that the impacts found in this study necessarily mean that the program would work equally well in any school that adopts it. This is not really a problem for the EAST program, since it has never been the intention of the developers that it should be required, but it must be emphasized that these findings should not be taken as evidence that districts should begin requiring (or necessarily even encouraging) schools to enroll in EAST if they are not inherently motivated to do so.

It should also be noted that all of the outcomes for which program impact was observed in this study were measured through self-report instruments. This is not a limitation for the motivation scales, since self-report is an appropriate way to measure attitudes, but it is perhaps a minor concern for the problem solving domains, which might be interpreted as measuring awareness rather than skills, since the instrument was not performance based. However, the instruments that were used for these measures were carefully researched by their developers and demonstrated strong psychometric properties in the development studies described by the authors as well as during the current study.

Outcomes that were measured on the student survey, howeverô particularly learning style, which was found to be affected by EAST participationô were derived from an unvalidated instrument, in addition to being measured through self-report. For this reason, among the five domains found to be affected by EAST, conclusions about learning style may be the least reliable. Nevertheless, the result was highly consistent with comments that were received during focus groups with EAST students. These students were very enthusiastic about their EAST classes and their facilitatorsø instructional methods, indicating that EAST made them more õeager,ö õconfident,ö and õindependentö than their regular classes did. They felt that because it is not õall laid out for youöô even in some cases with complex and frustrating tasksô it was helping prepare them for the real world, where their bosses wonøt tell them what to do. By making one think for oneself, developing research skills, and providing freedom to explore oneøs interests, they said, it made them more independent and effective in solving problems, and helped them to explore things in more depth.

Recommendations

Following are recommendations that we have developed as a result of this study. These include recommendations that we believe can help further maximize the success of this unique

program, as well as recommendations for researchers wishing to implement similar evaluation studies.

Recommendations for the program.

- The successes observed in this study provide additional reason to continue to expand the program, especially for the purpose of achieving the specific objectives that were identified as being affected by EAST. However, since participating schools were self-selected (and may have been even more motivated to participate than the typical EAST school, since their participation in EAST also required participation in the study), the study sample cannot be assumed to have been representative of schools state-wide. For this reason, these successes should not be taken as evidence that the program should be encouraged for any school who does not really want it.
- Recruitment methods were not clearly defined and had ambiguous results. While the state may not want to dictate such policies to the districts or schools, the importance of more proactively reaching out to populations who are less likely to seek out the program on their own could be stressed more strongly during training, with an explicit emphasis that simply casting a wide net is not sufficient to fully ensure diversity. Specifically, as they become better established, the programs might be encouraged to reach out more to traditionally underserved or under-represented groups, especially minorities, girls, and (to the extent appropriate for the program) special education students.
- Despite the fact that the observation rating form, while successful in assessing overall program fidelity, had limited success in evaluating the benefits of specific program components, it nevertheless possesses a strong conceptual alignment with program goals and philosophies. Just as EAST staff have already been using the logic model to inform their professional development process, the observation rating form could serve as a useful tool for focusing discussions around õsite healthö visits.
- Further study of program impacts could be valuable for helping to identify whether the skills and attitude changes that EAST develops have long-term impacts on studentsø further education, including post-high school graduation, or even choice of career paths. Such long-term effects have often been observed anecdotally, but have not been established methodically as a consistent program impact.
- Further study may also be warranted in order to more closely examine the extent to which program benefits may be influenced by other affective student characteristics, and the extent to which the program may be able to help students overcome certain characteristics such as a lack of interest in school. This could help further strengthen the value of the program by achieving a more targeted diversity, identifying and encouraging participation from types of students who might be shown to benefit but who traditionally might not apply. Since there are likely to be a number of student characteristics that are relevant to success other than the demographic indicators available from school records, school personnel who know the students may be in the best position to identify appropriate candidates. The targeted recruitment efforts that were often used by the schools can therefore be an effective way to accomplish this balance, as long as referring staff use appropriate criteria. We agree that diversity is a laudable goal for EAST enrollments, but it is not clear whether this should necessarily mean that enrollment should be representative of the school in all cases.

• Another potential objective for further research would be to explore the impacts of better established programs. It is quite possible that programs that have had the opportunity to become fully established might demonstrate even stronger impacts for a larger number of outcomes.

Recommendations for replication of the evaluation methodology.

- The value of development of a logic model early in the study cannot be overemphasized. Particularly for a program such as EAST, for which adaptability and responsiveness to specific circumstances is among its most important distinguishing features, it is critical to clearly define the parameters within which this flexibility is meant to exist, both to help new practitioners understand what is expected of them, and to help researchers recognize the program that is being evaluated. This process can be quite time consuming, particularly if the program design has never been explicitly laid out before, but it is in this very situation when it is needed the most. It should be noted, however, that it may not be possible to create a written program sufficiently to implement it or to evaluate it. It may very well be that a program such as EAST requires personal training from existing staff to be replicated or evaluated. This does not make the model any less real, but it does have implications for the cost of replication.
- Measuring program fidelity is a crucial part of program evaluation, especially when the programs that are being studied are in the earlier stages of implementation. Observation is an important method for collecting evidence of fidelity, especially when interactions between students and teachers form an essential part of the model. However, there may also be critical aspects of program fidelity that can not be assessed through passive observation alone, especially if resources do not allow for frequent enough observations to obtain a representative picture of the program in a short period of time. In such cases, interviews and/or surveys of site-based program staff and of monitoring staff can provide an invaluable complement to observation, and can also access potentially important factors (such as teachersø attitudes and beliefs) that would be extremely difficult if not impossible to infer from observation.
- Assessing the possibility of the existence of selection effects can be exceedingly difficult. It may be worth implementing additional behavioral or attitudinal measures to try to evaluate student characteristics that are expected to influence program outcomes, even if they do not represent goals of the program per se. However, since it is presumably impossible to directly assess all potentially salient variables, it is equally important to make inferences about the likelihood of unmeasured selection effects by examining the procedures used to identify participants. Of course, the best way to address this problem would be to use a randomized design, as was done for the EAST study. However, randomization of schools only addresses selection effects at the school level; selection effects at the student level would only be eliminated by a randomized design if selection or at least assignment can be randomized at the student level.
- Conducting a randomized evaluation design in a public school setting is extremely challenging, and in some circumstances, its pitfalls can be so severe that it may not be the best approach. Where the conditions exist that make it possibleô most notably, the opportunity to randomly select schools (or classrooms) for participation, or at least to

randomly assign applicants to participating and control conditionsô there are a number of additional conditions that must be achieved in order to maintain the integrity of the study. One of the most important issues is to ensure the cooperation of participating schoolsô including the control schoolsô throughout the study. Incentives can help but may not be sufficient if the schools do not perceive inherent value in a rigorous impact evaluation. Involving the schools in discussions about this value, or even in the process of designing the evaluation, can help, and, of course, providing schools with formative feedback throughout the study makes the evaluation more directly valuable to them. Involvement and sanctioning of the evaluation from authorities can also help, and laying out the obligations of participation in the evaluation as a condition for program participation can help ensure that schools stay the course. Nevertheless, it is important to avoid making schools feel like the study is being forced down their throats, as lack of cooperation can completely invalidate the randomization process. This is most challenging when control schools are being asked to participate actively in the study. Unless the program is already considered desirable (as was the case with EAST) and can be offered as a reward for cooperation (e.g., through delayed implementation), it might be best to design the study in such a way that only the most minimal involvement is required from control schools.

Sustainability, Capacity Building and Dissemination

Through its experience with the three-year evaluation of EAST, the Arkansas Department of Education has developed capacity to help it continue to use aspects of the program evaluation, and many of these tools can be applied to the evaluation of other educational technology initiatives. These tools include the logic model and the classroom observation protocols, which many EAST and ADE staff are now trained to use. In addition, many of these same staff worked closely with the evaluator in developing these tools and should therefore be better prepared to collaborate on similar efforts in the future. EAST staff also now have at their disposal a number of assessment instruments that could continue to be adapted and used for program evaluation, including the student survey, principal and facilitator surveys, end-of-year program rating forms, supplemental recruitment survey, student and facilitator focus group protocols, and (with permission from the authors) the motivation and problem solving inventories.

In addition to the recommendations presented in this report, the ADE will also work to assist other state education departments and local school districts in designing and carrying out evaluations of educational technology interventions by disseminating the results of this study. Many of the methods, practices analyses, and tools that were utilized for the EAST evaluation can be adapted for other studies. Among the most significant of these are the following:

- Development of a logic model in order to define the program characteristics in terms that are observable and measurable.
- Use of the logic model to develop observation protocols or other fidelity measures.
- Training observers in the use of the protocol, including the use of an annotated version to further define/clarify the elements of the logic model, and a set of guidelines to remind observers of protocols while they are in the field.
- Establishment of interrater reliability through group observations of classrooms not involved in the study.

- Re-validating measurement instruments on the target population by running confirmatory factor analysis.
- Addressing colinearity among subjects when participants are randomized at the group level through hierarchical linear modeling.

In order to facilitate ADE¢s efforts to disseminate these methods and insights, Metis Associates will create a user-friendly compact disc that summarizes the study methods and findings in terms that can be understood by a lay audience. This resource can be used for presentations which EAST and/or ADE staff may conduct at conferences such as the State Educational Technology Directors Association (SETDA) and the National Educational Computing Conference (NECC), but it will also be designed to stand on its own so that it can be posted, in whole or in part, on the EAST website and/or distributed through other venues.

In addition, Metis Associates will seek to disseminate the more technical aspects of the evaluation through presentations at professional conferences such as the American Evaluation Association and the American Educational Research Association. We are also exploring the possibility of publication of the study results in a peer review journal.

Bibliography

Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judge (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138-159). New York: Cambridge University Press.

Bloom, B (1984). Taxonomy of educational objectives. Boston, MA: Pearson Education. U.S.

Brookhart, Susan M. Review of the Iowa Tests of Basic Skills, forms K, L, and M. In Murphy, Linda L. (Ed.). (1998). *The thirteenth mental measurements yearbook*. Lincoln, Nebraska: The Buros Institute of Mental Measurements.

CDW-G. *Teachers Talk Technology 2005*. Retrieved August 8, 2007 from http://newsroom.cdwg.com/features/feature-08-29-05.htm.

Cohen, J. (1988). *Statistical power analysis for the social sciences (2nd ed.)*. New Jersey: Lawrence Erlbaum Associates.

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Educational and Psychological Measurement*. 20, 37-46.

Department of Education, Office of Educational Technology, Toward A New Golden Age in American Education: How the Internet, the Law and Today's Students Are Revolutionizing Expectations, Washington, D.C., 2004.

Dewey, M. E. (1983). Coefficients of agreement. British Journal of Psychiatry. 143, 487-489.

Drucker, P (1994). The age of social transformation. *The Atlantic Monthly*, 274 (5), 53-80. Retrieved October 17, 2007, from http://www.providersedge.com/docs/leadership_articles/Age_of_Social_Transformation.pdf

Frauenknecht, Marianne, & Black, David R. (1995). Social Problem-Solving Inventory for Adolescents (SPSI-A): Development and preliminary psychometric evaluation. *Journal of Personality Assessment*. 64(3), 522-539.

Frauenknecht, Marianne, & Black, David R. (2003). *The Social Problem-Solving Inventory for Adolescents (SPSI-A): A manual for application, interpretation, and psychometric evaluation.* Morgantown West Virginia: PNG Publications.

Hoover, H.D., Dunbar, S.B., & Frisbie, D.A. (2001). *Iowa Tests of Basic Skills complete/core battery: Norms and score conversion*. University of Iowa: Riverside Publishing.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. *33*, 159-174.

Lemke, C., & Coughlin, E. C. (1998). *Technology in American schools: Seven dimensions for gauging progress. A policymaker's guide*. The Milken Exchange on Educational Technology. Retrieved October 17, 2007, from http://www.mff.org/publications/publications.taf?page=158

McInerney, Dennis M. (2004). *Inventory of School Motivation: Guideline for use and validation checks*. Unpublished manuscript.

McInerney, Dennis M., Yeung, Alexander Seeshing, & McInerney, Valentina (2001). Crosscultural validation of the Inventory of School Motivation (ISM): Motivation Orientations of Navajo and Anglo Students. *Journal of Applied Measurment*. 2(2), 135-153.

McInerney, Dennis M. (1998, April). *Multidimensional aspects of motivation in cross-cultural settings and ways of researching this.* Paper presented at the American Educational Research Association Annual Meeting, San Diego, CA.

McInerney, Dennis M., Roche, Lawrence Al., McInerney, Valentina, & Marsh, Herbert W. (1997). Cultural perspectives on school motivation: The relevance and application of goal theory. *American Educational Research Journal*, *34*(*1*), 207-236.

McNabb, M., Hawkes, M., & Rouk, U. (1999, July). Critical issues in evaluating the effectiveness of technology. Paper presented at the National Conference on Educational Technology, Washington, DC.

Mehrens, William A. Review of the Iowa Tests of Educational Development, forms K, L, and M. In Murphy, Linda L. (Ed.). (1998). *The thirteenth mental measurements yearbook*. Lincoln, Nebraska: The Buros Institute of Mental Measurements.

Murray, J. (2003) Contemporary literacy: Essential skills for the 21st century. *MultiMedia Schools, vol. 10 (2),* March-April 2003. Retrieved July 21, 2003, from <u>http://www.infortoday.com/MMschools/mar03/murray.shmtl</u>.

The National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform (GPO Publication No. 065-000-00177-2). Washington, DC: U.S. Government Printing Office.

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

Reeves, T.C. (1998). The impact of media and technology in schools: A research report prepared for The Bertelsmann Foundation. Retrieved October 17, 2007, from http://it.coe.uga.edu/~treeves/edit6900/BertelsmannReeves98.pdf.

Ringstaff, C. and Kelley, L. (2002). *The Learning Return on Our Educational Technology Investment: A Review of Findings from Research*. WestEd Regional Technology in Education Consortium. Sackett, P. R., Laczo, R. M., & Arvey, R.D. (2002). The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. *Personnel Psychology*. 55, 807-825.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved September 7, 2005, from http://PAREonline.net/getvn.asp?v=9&n=4.

Subkoviak, Michael J. Review of the Iowa Tests of Educational Development, forms K, L, and M. In Murphy, Linda L. (Ed.). (1998). *The thirteenth mental measurements yearbook*. Lincoln, Nebraska: The Buros Institute of Mental Measurements.

Thornburg, D. (2005). *Why EAST Matters: EAST - Preparing Students for the Future*. The Thornburg Center. Retrieved October 17, 2007, from <u>http://www.tcpdpodcast.org/briefings/EAST.pdf</u>.

Appendices