Evaluation of a Multi-School Pilot Project Designed to Close Achievement Gaps

Georgia K. Hughes
Kimberly S. Cowley
Lisa D. Copley
Nicole L. Finch
Merrill L. Meehan

A paper presented at
2005 National Evaluation Institute
July 7 – 9, 2005
Memphis, TN

AEL
P.O. Box 1348
Charleston, WV 25325

.

Evaluation of a Multi-School Pilot Project Designed to Close Achievement Gaps

Georgia K. Hughes, AEL
Kimberly S. Cowley, AEL
Lisa D. Copley, AEL
Nicole L. Finch, AEL
Merrill L. Meehan, AEL

A paper presented at
2005 National Evaluation Institute
July 7-9, 2005
Memphis, TN

AEL
P. O. Box 1348
Charleston, WV 25325

Founded in 1966 as a not-for-profit corporation, AEL provides services to educators, education publishers, and policymakers. Services include rigorous research design and implementation, research reviews, intensive product and program evaluation, randomized field trials, technical assistance, and award-winning professional development programs. AEL operates several contracts funded by the U.S. Department of Education: a Regional Educational Laboratory, the Region IV Comprehensive Center, and an Eisenhower Regional Consortium for Mathematics and Science Education.

To contact AEL about research, products, or services:
P. O. Box 1348, Charleston, WV 25325; 304-347-0400 or 800-624-9120
info@ael.org • www.ael.org

AEL maintains policies and procedures that govern the Institutional Review Board in ensuring the welfare and protecting the rights of human subjects who choose to participate in research and evaluation activities.

This publication is based on work sponsored wholly or in part by the Institute of Education Sciences (IES), U.S. Department of Education, under contract number ED-01-CO-0016. Its contents do not necessarily reflect the positions or policies of IES, the Department, or any other agency of the U.S. government.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The achievement gap has been causing concern in education circles for many years. The term refers to the differences in academic achievement among ethnic and socioeconomic groups. An achievement gap often is seen between upper- and lower-class students and between students of differing races and ethnic backgrounds[1]. The consensus among researchers is that race and class are two major contributors to the achievement gap, among several others.

Although West Virginia's African American student population is small, consisting of approximately 4% of the state's 3rd-through 11th -grade population, a disproportionate percentage of African American students score in the lower quartiles on standardized tests. According to 2001-2002 data from the No Child Left Behind–West Virginia Report Cards (http://wveis.k12.wv.us/nclb/pub/pickinfo.asp), only 39% of the state's African American students in Grades 3 through 11 scored above the 50th percentile in basic skills on the Stanford 9. Thus, West Virginia, like other states across the nation, is not exempt from the challenges associated with improving the achievement of African American children and youth.

Background

Following the passage of the No Child Left Behind Act of 2001, many states and districts increased efforts to close the achievement gap. However, prior to the passage of that legislation, the seeds for a project focusing on improving the achievement of African American students found fertile ground in Kanawha County, which has one of the largest concentrations of African American student and adult populations in West Virginia. African Americans make up about 10% of the student population in Kanawha County Schools (KCS), Grades 3 through 11.

In 2001, KCS district officials asked an AEL staff member for help with planning ways to improve the academic achievement of African American students in the county. That initial request ultimately resulted in a pilot school project in which AEL worked with four schools and district leaders to improve instruction for all students, particularly those who are African American. In the fall of 2001, the KCS management team identified four schools in which to pilot efforts to improve academic achievement. The schools (two elementary, one middle, one high school) were selected based on their demographic composition (i.e., they had high percentages of African American students and students of low SES; see Table 1) and comprised a feeder pattern. During the 2001-2002 and 2002-2003 school years, AEL staff members established the necessary relationships and credibility with Pilot School staff and administrators to lay the groundwork for the full-scale pilot project and research study that were implemented in the 2003-2004 school year.

---

[1] In this report the term African American refers to students who are of African American descent. The term Black incorporates many different ethnic groups, e.g., Jamaican, Haitian, or African immigrants. Therefore, ethnic designations are used in their appropriate context.

Table 1
Selected Demographics of Pilot and Comparison Schools

School	2002-2003 School Demographics*		
	Total Students	Percentage of African American students	Free or Reduced-Price Lunch Percentage
Pilot Schools			
Elementary A	206	76%	78%
Elementary B	257	34%	67%
Middle School C	668	35%	60%
High School D	1,420	27%	33%
Comparison Schools			
Elementary E	358	33%	60%
Elementary F	226	59%	85%
Middle School G	464	30%	55%
High School H	1,087	18%	28%

*School demographic data are drawn from the common core of data; 2002-2003 is the most recent year for which data are available.

Research Purpose

The purpose of this project was to study instructional practices useful in improving the academic achievement of all students, particularly African American students. AEL staff believed that the approach of working closely with school and district personnel to provide professional development and exemplary culturally responsive instructional materials in selected Kanawha County schools would yield differentiated teacher and student classroom behaviors. Further, AEL believed that this approach would build the schools' capacity to continue improvement efforts aimed at raising the academic achievement of African American and other students. Thus, the research aimed to examine the impact of different levels of the intervention on teacher and student behaviors.

Research questions. The Pilot Schools project suggested several key questions around which research efforts were centered.

1. What is the effect on student achievement of schools' participation in a pilot research and development project designed to be responsive to the cultural needs of African American and low-socioeconomic status (SES) students?

2. In what ways, if at all, does schools' participation in a pilot research and development project to improve the achievement of children, especially African American and low-SES students, affect school staffs' sense of themselves as a high-performing learning community?

3. In what ways, if at all, does schools' participation in a pilot research and development project to improve the achievement of children, especially African American and low-SES students, affect schools' capacity to undertake improvement initiatives?

4. In what ways, if at all, does schools' participation in a pilot research and development project to improve the achievement of children, especially African American and low-SES students, alter students' perceptions of their schools' and families' support for students' academic endeavors and school climate conducive to learning?

5. To what degree does a teacher's participation in a full- or partial-treatment group affect the format, substance, and quality of his or her instruction?

6. To what degree does having a teacher in a full-treatment or partial-treatment group affect student engagement?

7. Over the course of this project, how do educators' interpretations of the achievement gap change?

Intervention Description

Culturally Responsive Teaching

The Pilot Schools project focused on culturally responsive teaching. Culturally responsive teaching is based on the idea that culture is central to student learning. This sociocultural approach to teaching, based on the work of Russian psychologist Lev Vygotsky, provides instructional scaffolding that encourages students to learn by building on the experiences, knowledge, and skills they bring to the classroom. Culturally responsive teaching is guided by research-based principles that address content, pedagogy, and teacher-student relationships. The Knowledge Loom, a Web site developed by the Education Alliance at Brown University (www.knowledgeloom.org), describes the following nine principles that were used to guide the Pilot Schools intervention.

- Communication of high expectations
- Active teaching methods
- Teacher as facilitator
- Positive perspectives on parents and families of culturally and linguistically diverse students
- Cultural sensitivity

- Reshaping the curriculum
- Culturally mediated instruction
- Student-controlled classroom discourse
- Small-group instruction and academically related discourse

Target Populations

Selected teachers at the four pilot schools in Kanawha County (two elementary schools, one middle school, and one high school) formed the target population for this intervention. The high school volunteered to participate; the elementary schools and the middle school were selected by the county superintendent to participate. All four pilot schools had African American student populations of at least 25%. Teachers in the pilot schools were placed into either the full treatment group or the partial treatment group. The *full treatment* group consisted of pilot team teachers at the pilot schools. These pilot school teams were composed of teachers who volunteered or were selected to participate in the project by their principal, and they ranged in size from 3 to10 members per school. The *partial treatment group* consisted of the remaining teachers from each of the pilot schools. The full and partial treatment groups were then subdivided on the basis of whether teachers taught a special culturally responsive curriculum unit.

In addition to the four pilot schools, four additional schools (two elementary schools, one middle school, and one high school) were selected by the superintendent to serve as comparison schools. These schools were selected because their demographic characteristics and achievement levels matched those of the pilot schools most closely, given the confines of Kanawha County (see Table 1 above). Teachers in the comparison schools school received no treatment.

Table 2 provides brief descriptions of the three intervention levels: (1) full-treatment groups, (2) partial-treatment groups, and (3) no-treatment comparison group.

Full Treatment

During the 2003-2004 school year, the full treatment group (Pilot School teams) participated in professional development sessions to learn about and enhance their skills related to culturally responsive instruction and in bimonthly team meetings to reinforce learning and facilitate reflection, sharing, and peer support. Pilot schools also received site-specific, ongoing technical assistance tailored to the requests of administrators or faculty. Additionally, some full treatment staff taught culturally responsive curriculum units codeveloped by AEL and KCS staff to demonstrate the principles of culturally responsive instruction. Thus, the full treatment group consisted of two subgroups: PTT (those who received the principles of training in culturally responsive instruction and

taught culturally responsive curriculum units) and PTNT (those who received the training but did not teach culturally responsive curriculum units).

Table 2
Pilot Schools Teacher Group Designations

Main Groups/Subgroups	Designation	Definition
(1) Full Treatment		
Pilot Team Teaching	PTT	Full treatment and taught CRU*
Pilot Team Not Teaching	PTNT	Full treatment and did not teach CRU
(2) Partial Treatment		
Non-Pilot Team Teaching	NPTT	Partial treatment and taught CRU
Non-Pilot Team Not Teaching	NPTNT	Partial treatment and did not teach CRU
(3) No Treatment		
Comparison schools	Comp.	No treatment and did not teach CRU

* CRU – Culturally Responsive Unit

Partial Treatment

The partial treatment group was composed of faculty members at each pilot school who were *not* serving on the Pilot School team. The partial treatment group did not participate in bimonthly meetings, and they did not have access to the technical assistance provided by the AEL facilitators. Some partial treatment staff taught the culturally responsive curriculum units provided by AEL. Thus, the partial treatment group consisted of two subgroups: NPTT (those who taught culturally responsive curriculum units) and NPTNT (those who did not teach culturally responsive instructional units).

No Treatment

The four comparison schools (Comp.) included two elementary schools, one middle school, and one high school (see Table 1 above). These schools and the teachers within the schools received no treatment and did not participate in any aspect of the intervention.

Table 3 presents the number of teachers in each group and details which components of the treatment they received. Participants in group PTT received the highest level of treatment, and participants in group NPTNT received the lowest level of treatment.

Table 3
Subgroup Numbers and Intervention Components Received

Group	N	Workshops or P.D.[a]	Taught CRU[b]	Bimonthly Meetings	Technical Assistance
Full Treatment (Pilot Teams)					
PTT	6	Yes	Yes	Yes	Yes
PTNT	16	Yes	No	Yes	Yes
Partial Treatment (Non-Pilot Teams)					
NPTT	18	Yes	Yes	No	No
NPTNT	140	Yes	No	No	No
No Treatment (Comparison)					
Comp.	128	No	No	No	No

a. Full Treatment participants attended training workshops. Partial Treatment participants received on-site professional development sessions.
b. CRU = culturally responsive unit

METHODS

Research Designs

The research design for the Pilot Schools project was a series of quasi-experiments, with two composed of only treatment and nontreatment groups, and with one design including a total of five groups: full treatment (subgroups PTT and PTNT), partial treatment (subgroups NPTT and NPTNT), and no treatment (Comp.). These quasi-experimental designs allowed some measure of statistical certainty that outcomes are due to pilot school activities rather than other education initiatives to which participants might have been exposed. These designs include three subdesigns: two for the teachers involved in this project and one for the students.

Teacher Designs

The first teacher design was based on dependent attitudinal variables and is labeled as "Untreated matched controls with pretests and posttests and various treatments" (Shadish, Cook, & Campbell, 2002, pgs. 153, 157). This design is diagrammed as

NR	0_1	X_1	X_2	X_3	0_2
NR					0_2

This design was used for the teacher-completed, paper-and-pencil AEL instruments to measure staffs' readiness for improvement and the extent to which the faculties were committed to continuous learning and improvement. The various treatment elements in the design refer to three different types of treatment available to them: external facilitation, internal facilitation, and implementation of culturally responsive curriculum units. This is a two-group design: treatment and control groups. A variation of this design for one instrument did not include a pretest.

The second teacher design was based on dependent behavioral variables and is labeled as "Untreated matched controls with dependent samples and various treatments" (Shadish, Cook, & Campbell, 2002, pgs. 153, 157). This design is diagrammed as:

						O_2
NR	O_1	X_1	X_2	X_3	O_3	
NR	O_1		X_1	X_2	O_3	
NR	O_1		X_2	X_3	O_3	
NR	O_1		X_2		O_3	
NR	O_1				O_3	O_2

This design was employed for the systematic observation of teachers' classrooms, hence the collection of dependent behavioral variables. This was a five-group design. Two groups were subgroups of the pilot teams (X_1), one subgroup that did teach the culturally responsive curriculum unit ($X_1 + X_2 + X_3$) and another subgroup that did all except teach the unit ($X_1 + X_2$). Two groups were teachers in the pilot school, but not on the pilot team (X_2); one subgroup received internal facilitation and did teach a unit ($X_2 + X_3$), and one subgroup did not teach the unit (X_2). It should be noted that systematic observations were completed in the classrooms of all five subgroups before, during, and after the culturally responsive curriculum units were taught by the two groups with X_3s. The arrow between the two O_2s conveys that the "during" observations occurred in the classrooms of all five groups.

Student Design

The student research design was based on the dependent attitudinal variables and student achievement data and is labeled "Untreated matched controls with a pretest and posttest and various treatments" (Shadish, Cook, & Campbell, 2002, pgs. 153, 157). The diagram for this design is the same as for the teacher attitudinal variables:

NR	0_1	X_1	X_2	X_3	0_2
NR					0_2

This design was used for the student-completed, paper-and-pencil AEL instrument to measure students' sense of belongingness, academic efficacy, and family expectations and involvement in education. Similar to the teacher designs, the various treatments refer to the different types of intervention components their classroom teachers received. This is because these students were in classrooms taught by teachers who received different treatment components. A variation of this design, which included no pretest, was used for the student achievement data.

Contextual Data Collection

Additionally, focus group and individual interviews with project participants provided context for quantitative findings and added richness and depth to the quasi-experimental designs. Focus groups were conducted with pilot team teachers at the beginning of the project to probe for their explanations of the origin of African American student academic underachievement. The Interview Design process for collecting similar qualitative data from the Pilot Schools project teachers was conducted at the end of the school year. Trained researchers conducted all qualitative research efforts using standardized interview protocols to ensure that data were collected systematically.

Instruments and Data Collection Protocols

AEL Continuous School Improvement Questionnaire

The AEL Continuous School Improvement Questionnaire (AEL CSIQ) is a 60-item instrument that measures a faculty's commitment to continuous learning and improvement. The names of the six, 10-item scales are provided below.

- Learning Culture
- School/Family/Community Connections
- Shared Leadership
- Shared Goals for Learning
- Purposeful Student Assessment
- Effective Teaching

The AEL CSIQ has demonstrated a high level of internal consistency reliability, with a Cronbach's alpha of .98 for the full instrument and alphas ranging from .91 to .96 for the six subscales (Meehan, Cowley, Craig, Balow, & Childers, 2002). The instrument has shown considerable stability over time, with test-retest correlations of .80 for the full instrument and range from .66 to .81 for the subscales. Concurrent validity and construct validity for the AEL CSIQ have also been established (see Meehan, et al.).

AEL Measure of School Capacity for Improvement

The AEL Measure of School Capacity for Improvement (AEL MSCI) assesses the degree to which schools possess the potential to become high-performing learning communities. The AEL MSCI consists of 64 items and eight subscales, named below.

- Collective Professional Capacity
- Peer-Reviewed Practice
- Program Coherence
- Technical Resources
- Anti-Discriminatory Teaching
- Responsive Pedagogy
- Differentiated Instruction

- Expectations for Student Performance

The AEL MSCI has demonstrated sufficient internal consistency reliability in the past, with a Cronbach's alpha of .97 for the full instrument and alphas ranging from .79 to .91 for the subscales (Riffle, Howley, & Ermolov, 2004). Test-retest reliability for the AEL MSCI is high, with a total-instrument test-retest correlation of .87 and correlations ranging from .68 to .86 for the eight subscales.

Special Strategies Observation System[2]

AEL staff used the Special Strategies Observation System (SSOS) to collect classroom behavior data for the pilot and comparison schools three times during the 2003-2004 school year (before, during, and after the culturally responsive units were taught in the pilot schools). This system, composed of three instruments, is designed to collect data systematically on essential elements of classroom behaviors related to instruction, management, and context. The three instruments include the Classroom Observation Form, QAIT Assessment of Classroom, and Classroom Environment and Resources Checklist. The entire observation could last a maximum of 60 minutes. Figure 1 portrays the SSOS system as a clock, with the time allotments specified for each instrument. The three instruments are described below.

Classroom Observation Form (COF). The COF is a combination observation system that is best described as a category system with low-inference items and multiple coding procedures (Nesselrodt & Schaffer, 1993; Sullivan & Meehan, 1983). It is based on the Stallings Observation System (Stallings, 1980) and the Classroom Activity Record designed by Evertson and Burry (1989). The top page of the COF collects typical demographic information, including the school, observer, date, teacher observed, number of adults and students in the class, grade level, subject being observed (teacher or target student), observation period (before, during, and after), and type of classroom (PTT, PTNT, NPTT, NPTNT, and Comp.).

The COF segment of the observation includes a maximum of 58 minutes—2 minutes for coding the cover page and then 56 minutes for coding classroom behaviors. The 56 minutes are divided into seven 8-minute time periods; each 8-minute block is captured on a separate page. The first minute of each block focuses on the entire classroom and provides a class snapshot by looking at both student engagement (the number of students on task, off task, waiting, or out of the room) and groups and activities (whether students are clustered in teacher, aide, or student groups and their type of involvement, such as working alone, management, interaction, or socialization). The remainder of each 8-minute block is devoted to observing either the teacher or "target" student.

For the pilot schools project, researchers decided to include both the teacher observation and a target student (African American youth were selected as the target

[2] The SSOS was revised in 2004 and renamed as the Special Strategies Observation System-Revised (SSOS-R).

students for observed classrooms). The focus of the observation switched from teacher to target student for each 8-minute block. There were a total of 27 discrete activities that could be chosen to describe the teacher and target student behaviors.

Figure 1
Graphic Depiction of the Special Strategies Observation System (SSOS) and Time Allotted to the QAIT and CERC

QAIT Assessment of Classroom. This instrument is best described as a moderate- and high-inference, simple coding, rating device. QAIT stands for Quality of Instruction, Appropriate Level of Instruction, Incentives for Learning, and Use of Time. This two-page instrument contains 40 items grouped under those four major categories.

Each item has a Likert-type response scale of 1 to 5 (*unlike this class* to *like this class*). This instrument was completed at the end of each observation session.

Classroom Environment and Resources Checklist (CERC). This instrument is a low-inference, simple coding, sign system. This one-page checklist contains 14 classroom attributes that are coded either as present or not present, such as adequate lighting, use of multiracial materials, posted assignments, etc. Two items were added that were specific to the pilot schools project: culturally mediated instruction and student-controlled classroom discourse. Next, 18 classroom resource items, such as textbooks, computers, and worksheets, are listed. Observers indicate first whether such resources were visible or not. If visible, observers then indicate whether the resources were used during the observation. This instrument was completed at the end of each observation session.

The SSOS instruments possess face and content validity and have proven their utility in prior research (Nesselrodt & Schaffer, 2000a, 2000b; Cowley et al., 2002). A high degree of inter-rater reliability was achieved among the data collectors during the 2001 and 2003 SSOS training sessions. To assess the degree of internal consistency reliability, Cronbach Alpha coefficients were computed for this administration of the SSOS. For the COF instrument, the activity code section coefficient was .51, the student engagement coefficient was .76, and the grouping strategy coefficients were .76 for the grouping section and .38 for the number of students section. For the QAIT instrument, the coefficient for all 40 items was .96; the coefficients for each of the scales ranged from .69 to .95. For the CERC instrument, the coefficient for all 50 items was .87; the coefficients for each section ranged from .44 to .81.

AEL Measure of Academic Supportiveness and Climate

The AEL Measure of Academic Supportiveness and Climate (AEL MASC) is a student-completed, paper-and-pencil instrument. The 42-item instrument assesses students' perceptions of themselves as students and of their school experiences, and also asks students about their families' awareness of and involvement in their children's school lives. The four scales within the instrument are: Student Belonging, Family Expectations, Student Academic Efficacy, and Family/School/Student Involvement. Taken together, these subscales assess the degree to which students think their schools and families provide them with academic nurturance and support, and the extent to which students view themselves as intellectually capable.

The AEL MSCI has demonstrated a high level of internal consistency reliability, with Cronbach's alphas of .95 for the full instrument and alphas ranging from .78 to .93 for the four subscales. Test-retest analyses indicate that the instrument measures the constructs reliably over time; test-retest correlations have demonstrated stability for the full instrument over time (r = .68) and for the four factors (range from .52 to .76) (Cowley, Riffle, Howley, Voelkel, & Ermolov, 2004).

Student Achievement Assessment

The West Virginia Department of Education (WVDE) employs a customized test to assess student achievement (WVDE, 2004). The West Virginia Educational Standards Test (WESTEST) is a criterion-referenced test that is designed to align with the state's Content Standards and Objectives (CSOs). WESTEST is administered to all public school students in Grades 3 through 8 and Grade 10, and addresses several content areas: mathematics, reading/language arts, science, social studies (except for Grade 10 in which social studies is not assessed). Items on the test include multiple choice, short answer, and constructed response, and the items require students to apply various thinking skills (e.g., knowledge/recall/recognition, synthesis/evaluation/extended thinking). Student performance on WESTEST is measured by a scale score, which can then be translated into a performance level. The five performance levels (Novice, Partial Mastery, Mastery, Above Mastery, Distinguished) are based on specific criteria for each grade level and content area.

Interview Design Process

Interview Design is a data collection process in which all participants ask questions, answer questions, and analyze responses. Interview Design was employed during the spring 2002 training workshop to gather data concerning perceptions of teachers, administrators, and community members associated with the Pilot Schools project about several issues, including lessons learned about teaching African American students, decisions made outside the classroom that impact African American students, factors contributing to African American students' higher level work, factors contributing to the achievement gap, and instructional strategies that hold promise for narrowing the achievement gap.

Data Collection

AEL Continuous School Improvement Questionnaire

The AEL CSIQ instruments were administered to the professional staffs at pilot and comparison schools during late April and early May of 2004. AEL staff encouraged school staff to administer instruments to the professional staff in a group setting. However, some schools may have distributed the questionnaires to their professional staff individually. Completed questionnaires were returned to AEL's offices in early and mid-May of 2004. The data were then scanned, entered into databases, cleaned, and prepared for analysis.

AEL Measure of School Capacity for Improvement

The AEL MSCI also was administered to pilot school professional staff in the spring of 2003. The procedures and time frame for administering the AEL MSCI in the spring of 2004 were identical to the data collection procedures used for the AEL CSIQ. AEL MSCI instruments were received with the AEL CSIQ questionnaires in early and

mid-May of 2004. AEL MSCI data were scanned, entered into databases, cleaned, and prepared for analysis.

Special Strategies Observation System

Three pairs of data collectors were assigned to specific schools, but each data collector completed his/her classroom observations individually. All data collectors utilized the SSOS forms during the classroom observations. The COF instrument was fully completed while in the classroom. Given that some observations ran consecutively, it was not always possible for the data collectors to fully complete the QAIT and CERC while in the classroom, but these instruments were completed as soon after the observations as possible.

Observations took place during the 2003-2004 school year before the culturally responsive units were taught, while the units were being taught, and after the units were taught. Teachers were selected for participation in the observation process based on a number of criteria, including consideration of pilot school group membership (i.e., PTT, PTNT, NPTT, NPTNT; refer to Tables 2 and 3). After teachers in group PTT were identified, along with the grade level and subject in which the culturally responsive unit was being taught, then classrooms were selected across the other three pilot school groupings (i.e., PTNT, NPTT, NPTNT) and the comparison school group (Comp.) to match the grade levels and subject areas of teachers in group PTT.

Each data collector (or in some cases, a guidance counselor) chose a "target" student for each classroom he/she observed; target students were selected as randomly as possible. Given that the focus of the pilot schools project was on closing the achievement gap between White and minority youth, African American males and females were chosen as target students, when possible, for these observations.

The number of classrooms selected for observation across the eight schools ranged from five to nine. A total of 315 classroom observations were completed during the 2003-2004 school year. SSOS data were scanned by observation period for each of the eight schools; data files were then cleaned, exported to SPSS, and merged into one master file for statistical analysis.

AEL Measure of Academic Supportiveness and Climate

The AEL MASC was administered to students at the pilot and comparison schools in April and May of 2004. The questionnaire also was administered to pilot school students in the spring of 2003. School staff were asked to administer the instruments to all students during English and Language Arts classes; however, some schools may have administered the AEL MASC at different times (e.g., during enrichment time). School staff returned the completed questionnaires to AEL's offices during early and mid-May of 2004. The data were scanned, entered into databases, cleaned, and prepared for analysis.

Student Achievement Assessment

All public school students in West Virginia were given the WESTEST in March of 2004. The tests were sent to CTB/McGraw-Hill for scoring. KCS officials received students' scores in August of 2004, and AEL researchers were given access to those data in late August. In collaboration with the director of counseling and testing for Kanawha County, AEL research staff collected scale scores and performance-level data for students enrolled in classrooms observed for the Pilot Schools project during the course of the 2003-2004 school year. Each case was assigned a unique code number that could not be used to identify the student or link WESTEST data with the individual.

Interview Design Process

During the Interview Design process used at the spring 2004 workshop, participants sat at long tables with two lines of five on each side. Each person received a sheet of paper with one of five questions printed on it, with room to record responses. Participants asked their question to the person in the facing chair and recorded the respondent's answer. After 3 to 5 minutes, the respondents and interviewers switched roles, and those who had previously been respondents asked their questions of their former interviewer. After an additional 3 to 5 minutes, one row of people at each table, who were designated as "movers," were instructed to move one seat to the left, and the entire process was repeated five times until each person had responded to every question.

Data Analysis

AEL Continuous School Improvement Questionnaire

To compare the differences on the AEL CSIQ between the pilot schools and the comparison schools, independent *t* tests were computed on each subscale using an alpha level of .05. To compare differences between comparison and pilot schools by building level (elementary, middle, high), independent *t* tests also were conducted for each subscale. To compensate for the multiple comparison problem of increasing the Type I error when analyzing more than two groups, the Bonferroni procedure was used to adjust the original .05 alpha level. The revised alpha level for each building-level specific *t* test was set at .017. Effect sizes were computed as appropriate.

AEL Measure of School Capacity for Improvement

To compare the differences on the AEL MSCI between the pilot schools and the comparison schools, independent *t* tests were computed on each subscale using an alpha level of .05. To compare differences between pilot and comparison schools at the building level (elementary, middle, high), independent *t* tests were also conducted for each subscale. Again, the Bonferroni procedure was used to adjust the original .05 alpha level to compensate for the multiple comparison problem. The revised alpha level for each building-level specific *t* test was set at .017. Effect sizes were computed as appropriate.

To compare differences on the AEL MSCI between two administrations (spring 2003 and spring 2004) of the survey to pilot school staff, mean subscale scores and standard deviations were computed. In addition, independent *t* tests were computed using an alpha level of .05. Again, effect sizes were computed as appropriate.

Special Strategies Observation System

Descriptive (e.g., frequencies, percentages) and inferential (e.g., ANOVA) statistics were computed as appropriate using the classroom observation data.

COF. COF activity data provided the number of minutes spent in any of 27 discrete activities for both the teacher and the target student. The minutes were summed across the 8-minute intervals for each observation by both teacher and target student. These data were analyzed for each of the 27 individual activities and by collapsing the data into four main categories of teacher-led, student-led, management/organization, and off-task.

COF classroom snapshot data for student engagement (i.e., the number of students on task, off task, out of the room, or waiting during the first minute of each 8-minute block) were summed across the 8-minute intervals for each observation, by both teacher and target student, and also by determining the percentage of students engaged in each category (on task, off task, out of room, waiting). These data were analyzed using the four engagement codes of on or off task, out of the room, and waiting. For the groups and activities segment of the COF, the number of students involved with the teacher, aide, or other students by activity (interactive, working alone, management, or social/uninvolved) were summed across the 8-minute intervals for each observation by both teacher and target student. These data were analyzed by number of students per activity and also by determining the percentages of students engaged in each type of activity. Further, one-way ANOVAs were conducted for all of the COF data (activity, student engagement, and groups and activities) to determine if statistically significant differences existed among the five groups. Effect sizes were computed as appropriate.

QAIT. QAIT data were analyzed by creating four scales composed of the 40 individual items: quality of instruction, appropriate level of instruction, incentives for

learning, and use of time. Because there were unequal numbers of items in each scale, the item scores were summed and then averaged to generate the scale scores. Descriptive statistics were used to describe results for each of the five groups. Further, one-way ANOVAs were conducted to determine if statistically significant differences existed among the group scale scores. As appropriate for the analyses, effect sizes were also computed.

CERC. CERC data were analyzed by calculating frequency percentages showing whether the classroom attributes were present and whether the classroom resources were visible and used during the observations for each of the five groups.

Adherence Index. To determine the adherence to the principles of culturally responsive instruction, an index score was generated from those QAIT and CERC data points that most closely aligned to any of the nine specific components of culturally responsive instruction. See Table 10 for a depiction of the nine components, along with the associated QAIT and CERC items. The scores for these items were converted to *z* scores and then to a standardized *Z* score (multiplying the *z* score by 10 and adding 50 generates a mean of 50 with a standard deviation of 10). The mean score for each component's group of items was generated as a new variable and analyzed by grouping.

AEL Measure of Academic Supportiveness and Climate

To compare the differences on the AEL MASC between the pilot schools and the comparison schools, independent *t* tests were computed on each subscale using an alpha level of .05. To compare differences between comparison and pilot schools by building level (elementary, middle, high), independent *t* tests also were conducted for each subscale. As with the AEL CSIQ and the AEL MSCI analyses, the Bonferroni procedure was used to adjust the .05 alpha level to compensate for multiple comparison problems. The revised alpha level for each building-level specific *t* test was set at .017. Effect sizes for these analyses were computed.

To compare differences on the AEL MASC between two administrations (spring 2003 and spring 2004) of the survey to pilot school students, mean subscale scores and standard deviations were computed. In addition, independent *t* tests were computed using an alpha level of .05. As with other analyses conducted in this research, effect sizes were computed as well.

Student Achievement Assessment

Research staff examined WESTEST performance-level data and scale scores by grade level for three content areas (mathematics, reading/language arts, science) for pilot and comparison schools. Descriptive data (e.g., frequencies) were examined as appropriate. The percentages of students at or above mastery were calculated for each grade level in each of the three content areas. Further, AEL researchers performed independent group *t*-tests using the scale scores for each grade level in each of the content areas. Because there was some concern about unequal sample sizes, Levine's test

for equality of variances was performed, and the correction for unequal variances was applied as necessary.

Table 4
Culturally Responsive Unit Components and Aligned QAIT and CERC Items

Component	Instrument	Item
Communication of high expectations	QAIT	10g. Communicating high expectations.
Active teaching methods	QAIT	9a. Presenting surprising demonstrations. 9b. Relating topics to students' lives. 9c. Allowing students to discover information.
Teacher as facilitator	QAIT	1a. Organizes information in an orderly way. 1b. Notes transitions to new topics. 1d. Frequently restates essential principles. 3. The teacher exhibits enthusiasm. 6. Teachers use appropriate pace to cover content. 7a. Accommodates students' levels of knowledge. 7b. Accommodates students' learning rates. 13a. Necessary time is allocated for instruction. 14a. The teacher uses effective management.
Positive perspectives on parents and families of culturally and linguistically diverse students	QAIT	1c. Uses many vivid images and examples.
Cultural sensitivity	QAIT	2a. Uses devices such as advanced organizers. 2b. Reminds students of previously-learned mat. 9d. Presenting intrinsically interesting material.
Reshaping the curriculum	--	No items directly aligned to this component.
Culturally mediated instruction	CERC	Culturally mediated instruction environmental indicator.
Student-controlled classroom discourse	CERC	Student-controlled classroom discourse environmental indicator.
Small group instruction and academically related discourse	QAIT	8a. Uses in-class ability grouping. 8b. Has a class that is homogeneous in ability. 8c. Uses cooperative learning arrangements.

Interview Design Process

Responses to Interview Design questions were reviewed for overarching, repeating categories, which were assigned broad codes. Finer coding was employed to identify emerging patterns within each broad category. The responses were then analyzed by theme and tabulated to provide a general, quantitative analysis of the most salient and prevalent issues that arose during the discussions.

FINDINGS

AEL Continuous School Improvement Questionnaire

Each of the six AEL CSIQ subscales contains 10 items, which respondents rated using a scale of 1 to 6 (*not present* to *present to a high degree*). These ratings were added together to create subscale mean scores ranging from 10.00 to 60.00. Table 5 provides descriptive statistical summaries for the six subscales for the pilot and comparison schools and by building level for each of those groups.

In general, comparison schools had higher mean scores across the six scales than the pilot schools. When looking at the data for the full groups, statistically significant differences were found on three of the six subscales: School/Family/Community Connections, Shared Goals for Learning, and Effective Teaching. All three analyses had small effect sizes (Cohen *d* column on Table 5), indicating some practical meaningfulness in the differences.

Independent *t* tests also were computed comparing pilot schools and comparison schools at each building level on the six subscales. Statistically significant differences were found between elementary schools on four of the subscales (Shared Goals for Learning, Effective Teaching, Learning Culture, School/Family/Community Connections). At the middle school level, statistically significant differences were found on four of the subscales (Learning Culture, Shared Goals for Learning, Shared Leadership, School/Family/Community Connections). Statistically significant differences also were found on one subscale between high schools (Shared Leadership), for which the pilot school had a higher score than the comparison school. Medium to large effect sizes were found for each of these differences. See Table 5 for more details.

AEL Measure of School Capacity for Improvement

Each of the eight AEL MSCI subscales contains eight items, which respondents rated using a scale of 1 to 4 (*not at all/never true* to *almost always/frequently true*). The ratings for each subscale were added together and divided by eight to create a mean score for that subscale ranging from 1 to 4. Table 6 provides descriptive statistical summaries for the eight subscales for the pilot and comparison schools, as well as building-level scores for each of those groups.

In general, comparison schools had higher subscale mean scores across the eight scales than did the pilot schools. When looking at the data for the full group, statistically significant differences were found on four of the eight subscales: Collective Professional Capacity, Technical Resources, Differentiated Instruction, and Expectations for Student Performance. Small and moderate effect sizes were found for all of these differences.

Table 5
AEL CSIQ Subscale Descriptive Statistics and Differences by Full Group and Building Level

Subscale Name	School Level	Pilot			Comparison			*df*	*t*	*p*	Dif.	*d*
		N	*Mean*	*SD*	*N*	*Mean*	*SD*					
Learning Culture	Elementary	27	48.37	7.67	29	53.62	5.15	54.0	3.03	.004	-5.25	0.80
	Middle	45	44.44	7.84	27	49.48	6.73	70.0	2.78	.007	-5.04	0.69
	High	53	47.28	8.21	57	45.18	8.58	108.0	1.31	ns	2.11	0.25
	Full Group	125	46.50	8.06	113	48.37	8.16	236.0	1.78	ns	-1.88	0.23
School/Family/Community Connections	Elementary	27	46.63	9.32	31	53.97	4.59	36.8	3.72	.001	-7.34	1.00
	Middle	44	43.34	8.71	26	50.27	7.35	68.0	3.40	.001	-6.93	0.86
	High	50	45.22	8.73	57	44.40	8.74	105.0	0.48	ns	0.81	0.09
	Full Group	121	44.85	8.87	114	48.34	8.53	233.0	3.07	.002	-3.49	0.40
Shared Leadership	Elementary	28	49.46	8.89	32	53.56	7.06	58.0	1.99	ns	-4.10	0.51
	Middle	46	42.20	11.04	27	50.26	9.30	71.0	3.19	.002	-8.06	0.79
	High	51	45.75	8.80	60	38.92	10.24	109.0	3.73	.000	6.83	0.72
	Full Group	125	45.27	10.00	119	45.43	11.37	242.0	0.11	ns	-0.16	0.01
Shared Goals for Learning	Elementary	27	50.11	8.16	31	54.81	5.82	56.0	2.55	.014	-4.70	0.66
	Middle	43	43.07	8.54	26	49.50	8.43	67.0	3.05	.003	-6.43	0.75
	High	48	48.85	7.75	55	46.31	8.43	101.0	1.59	ns	2.55	0.31
	Full Group	118	47.03	8.62	112	49.40	8.52	228.0	2.09	.037	-2.37	0.28
Purposeful Student Assessment	Elementary	26	50.46	7.56	31	53.81	5.28	55.0	1.96	ns	-3.34	0.51
	Middle	43	44.63	7.84	27	48.59	9.37	68.0	1.91	ns	-3.96	0.46
	High	51	47.16	7.41	56	44.43	8.74	105.0	1.73	ns	2.73	0.34
	Full Group	120	46.97	7.84	114	47.96	8.97	232.0	0.91	ns	-1.00	0.12
Effective Teaching	Elementary	27	50.85	6.73	31	54.65	4.28	56.0	2.60	.012	-3.79	0.67
	Middle	44	46.14	8.73	27	50.44	6.51	69.0	2.21	ns	-4.30	0.56
	High	52	48.42	7.69	59	47.76	7.69	109.0	0.46	ns	0.66	0.09
	Full Group	123	48.14	8.02	117	50.21	7.11	238.0	2.11	.036	-2.07	0.27

Independent *t* tests also were computed comparing pilot schools and comparison schools at each building level on the eight subscales. Statistically significant differences were found between elementary schools on three of the subscales: Program Coherence, Anti-Discriminatory Teaching , and Technical Resources. At the middle school level, statistically significant differences were found on seven of the eight subscales (Peer Reviewed Practice being the exception). Large effect sizes were found for the significant elementary level and middle school level differences. Statistically significant differences with small effect sizes also were found on two subscales between high schools (Peer Reviewed Practice and Program Coherence). See Table 6 for more detail.

Table 7 provides descriptive statistical summaries for the eight subscales across two administrations of the AEL MSCI to pilot school professional staff. In general, the spring 2004 administration showed higher mean scores across the eight scales than did the spring 2003 administration. Independent *t* tests were computed comparing both administrations on the eight subscales. Statistically significant differences were found on six of the eight subscales: Peer-Reviewed Practice, Collective Professional Capacity, Responsive Pedagogy, Technical Resources, Program Coherence, and Anti-Discriminatory Teaching. Of these differences, the change in Anti-Discriminatory Teaching scores had the largest effect size (.97).

SSOS Classroom Observations

A total of 315 observations took place in 54 different classrooms in the eight schools (four pilot schools, four comparison schools) during the 2003-2004 school year. Data collectors tried to complete two observations per teacher at each of the three time periods throughout the year (before, during, and after teaching of the culturally responsive units). The number of observations per school ranged from 30 to 48 over the school year.

The SSOS allowed for a maximum of seven 8-minute observation blocks; four of these were focused on the teacher and three were focused on the target student. During the 8-minute blocks, 27 discrete activities could be coded in time segments of 1 to 8 minutes. These blocks equaled a maximum of 56 possible behavior coding minutes during the 60-minute observation; 2 minutes were dedicated to preliminary coding on the cover page at the beginning of the observation and 2 minutes each were allocated to completing the QAIT and CERC at the conclusion of the observation.

The average number of minutes of classroom coding per observation was 41.5 (standard deviation of 8.4). The classroom coding ranged from a low of 18 minutes to a high of the maximum 56 minutes. A total of 13,073 classroom behavior minutes were coded. Taking into consideration the 4 additional minutes per observation (for QAIT and CERC coding) and multiplying by 315 observations (1,260 minutes) brings the total number of observation minutes to 14,333 minutes or 239 hours (equivalent to roughly 30 days of observations).

Table 6
AEL MSCI Subscale Descriptive Statistics and Differences by Full Group and Building Level

Subscale Name	Level	Pilot Schools			Comparison Schools			*df*	*t*	*p*	Dif.	*d*
		N	Mean	*SD*	*N*	Mean	*SD*					
Collective Professional Capacity	Elementary	29	3.30	0.28	33	3.47	0.36	60	2.02	ns	-0.17	0.52
	Middle	49	2.88	0.44	29	3.41	0.37	76	5.48	.000	-0.53	1.31
	High	53	3.16	0.44	60	3.13	0.44	111	0.43	ns	0.04	0.08
	Full Group	131	3.09	0.44	122	3.28	0.43	251	3.63	.000	-0.20	0.46
Peer Reviewed Practice	Elementary	29	3.29	0.50	33	3.46	0.44	60	1.46	ns	-0.17	0.37
	Middle	49	2.88	0.59	29	3.17	0.49	76	2.17	ns	-0.28	0.52
	High	53	2.87	0.63	60	2.59	0.55	111	2.48	.015	0.28	0.47
	Full Group	131	2.97	0.61	122	2.96	0.63	251	0.04	ns	0.003	0.01
Program Coherence	Elementary	29	3.36	0.34	33	3.65	0.29	60	3.64	.001	-0.29	0.92
	Middle	49	2.81	0.46	29	3.18	0.37	76	3.66	.000	-0.37	0.88
	High	53	3.20	0.51	60	2.96	0.51	111	2.51	.014	0.24	0.47
	Full Group	131	3.09	0.51	122	3.20	0.52	251	1.67	ns	-0.11	0.21
Technical Resources	Elementary	29	2.85	0.48	33	3.38	0.38	60	4.84	.000	-0.53	1.22
	Middle	49	2.72	0.44	29	3.08	0.30	76	3.88	.000	-0.36	0.95
	High	53	2.85	0.43	60	2.88	0.45	111	0.47	ns	-0.04	0.09
	Full Group	131	2.80	0.45	122	3.07	0.45	251	4.70	.000	-0.27	0.59
Anti-Discriminatory Teaching	Elementary	29	3.75	0.22	33	3.92	0.14	45.53	3.55	.001	-0.17	0.92
	Middle	49	3.53	0.48	29	3.82	0.27	75.89	3.42	.001	-0.29	0.75
	High	52	3.67	0.38	60	3.50	0.43	110	2.14	ns	0.17	0.41
	Full Group	130	3.63	0.40	122	3.69	0.39	250	1.16	ns	-0.06	0.15
Responsive Pedagogy	Elementary	29	3.55	0.28	33	3.70	0.24	60	2.26	ns	-0.15	0.57
	Middle	49	3.23	0.52	29	3.63	0.29	75.80	4.38	.000	-0.40	0.95
	High	53	3.35	0.49	60	3.19	0.47	111	1.74	ns	0.16	0.33
	Full Group	131	3.35	0.48	122	3.43	0.45	251	1.43	ns	-0.08	0.18
Differentiated Instruction	Elementary	29	3.50	0.36	32	3.50	0.41	59	1.76	ns	-0.17	0.45
	Middle	49	3.05	0.50	29	3.52	0.47	76	4.12	.000	-0.47	0.97
	High	53	3.20	0.55	60	3.11	0.54	111	0.93	ns	0.10	0.17
	Full Group	131	3.21	0.52	121	3.36	0.55	250	2.14	.033	-0.14	0.27
Expectations for Student Performance	Elementary	29	3.46	0.44	32	3.32	0.50	59	1.12	ns	0.14	0.29
	Middle	49	2.80	0.47	29	3.47	0.38	76	6.47	.000	-0.67	1.56
	High	53	3.20	0.51	60	3.22	0.54	111	0.16	ns	-0.02	0.03
	Full Group	131	3.11	0.54	121	3.30	0.50	250	2.98	.003	-0.20	0.38

Table 7
AEL MSCI Subscale Descriptive Statistics and Differences for Pilot Schools by Year of Administration

Subscale Name	Spring 2003			Spring 2004			*df*	*t*	*P*	Dif.	*d*
	N	Mean	*SD*	*N*	Mean	*SD*					
Collective Professional Capacity	132	2.78	0.46	131	3.09	0.44	261	5.41	.000	0.30	0.67
Peer Reviewed Practice	130	2.12	0.84	131	2.42	0.94	259	2.75	.006	0.30	0.34
Program Coherence	132	2.87	0.40	131	3.09	0.51	246.94	3.88	.000	0.22	0.48
Technical Resources	132	2.60	0.59	131	2.91	0.52	261	4.57	.000	0.31	0.56
Anti-Discriminatory Teaching	131	3.24	0.40	130	3.63	0.40	259	7.87	.000	0.39	0.97
Responsive Pedagogy	131	3.09	0.42	131	3.35	0.48	260	4.72	.000	0.26	0.58
Differentiated Instruction	131	3.17	0.44	131	3.23	0.53	251.48	0.98	ns	0.06	0.12
Expectations for Student Performance	131	3.09	0.42	131	3.11	0.54	245.13	0.46	ns	0.03	0.06

Classroom snapshot: Student engagement. The four categories within student engagement included number of students on task, off task, out of the room, and waiting; these categories were further divided into those snapshots taken during a teacher-focused block or a target student-focused block. On a global level, the average number of students on task when the focus was the target student was 43, with a standard deviation (SD) of 15; for the teacher focus, the average was 49, with an SD of 18. For students off task, the overall average during the target student focus was 3 (SD = 5); for teacher focus, the average was 3 (SD = 6). For students out of the room, the target student-focused average was 3 (SD = 6); the teacher-focused average was 4 (SD = 7). Finally, for the number of students waiting, the target student-focused average was 2 (SD = 7); the teacher-focused average was 4 (SD = 9).

Table 8 provides descriptive statistics for the student engagement variables (on task, off task, out of room, and waiting) by target student or teacher focus and by grouping (PTT, PTNT, NPTT, NPTNT, Comp.). The overwhelming majority of students for both target student and teacher focus for all five groups were coded as on task.

One-way ANOVAs were generated to determine whether statistically significant differences existed within these eight variables (on task, off task, out of the room, and waiting by either target student or teacher focus) by grouping (PTT, PTNT, NPTT, NPTNT, Comp.). Five of the eight ANOVAs did identify significant differences by grouping, and an appropriate post hoc procedure (Dunnett's *C*) was selected to identify which groups were statistically significantly different from one another. For target student focus, the two categories of off task and out of the room resulted in statistically significant differences by group: for off task, PTT < PTNT, NPTT; for out of room, Comp. > PTT, PTNT, NPTT. For teacher focus, three categories of on task, off task, and out of the room were statistically significant by group: for on task, Comp. > NPTT; for off task, PTT < all four other groups; for out of room, Comp. > all four other groups. With one small effect size at .21 and the other four medium effect sizes above .25, these findings suggest that there was some practical significance. See Table 9 for a summary of these results.

Finally, the percent of students engaged in each category (on task, off task, out of the room, or waiting) was calculated by the five groupings (PTT, PTNT, NPTT, NPTNT, Comp.). According to Stallings (1980), student engagement rates above 80% have been associated with high gains in student achievement. For four of the five groups (PTT, PTNT, NPTNT, Comp.), the percentage of students on task was above 80%, and for the PTT group, the percentage exceeded 90%. For the NPTT group, the percentage approached 80%. See Table 10 for a summary of category percentages by grouping.

Table 8
Descriptive Statistics for Number of Students in SSOS Student Engagement by Target Student or Teacher Focus and by Grouping

Group	Statistic	Students On Task	Students Off Task	Students Out of Room	Students Waiting
Target Student Focus					
Pilot Teachers Teaching CRU*	*N* observed	33	33	33	33
	Mean	43	1	2	0
	SD	13	2	3	0
Pilot Teachers Not Teaching CRU	*N* observed	43	43	43	43
	Mean	45	4	1	1
	SD	16	6	2	4
Non-Pilot Teachers Teaching CRU	*N* observed	36	36	36	36
	Mean	40	6	2	3
	SD	9	9	3	6
Non-Pilot Teachers Not Teaching CRU	*N* observed	47	47	47	47
	Mean	41	3	2	4
	SD	16	5	5	11
Comparison Teachers	*N* observed	156	156	156	156
	Mean	45	2	5	2
	SD	15	4	7	6
Teacher Focus					
Pilot Teachers Teaching CRU*	*N* observed	33	33	33	33
	Mean	48	1	3	1
	SD	11	1	3	4
Pilot Teachers Not Teaching CRU	*N* observed	43	43	43	43
	Mean	48	6	2	5
	SD	20	9	2	10
Non-Pilot Teachers Teaching CRU	*N* observed	36	36	36	36
	Mean	40	6	2	5
	SD	13	10	3	8
Non-Pilot Teachers Not Teaching CRU	*N* observed	47	47	47	47
	Mean	47	3	3	4
	SD	14	5	5	8
Comparison Teachers	*N* observed	156	156	156	156
	Mean	52	3	6	5
	SD	19	4	9	10

*CRU = culturally responsive unit

Table 9
One-Way ANOVA Results for Number of Students in SSOS Student Engagement by Target Student or Teacher Focus

Engagement Category	*df*	*F*	*p*	Effect size	Difference
Target Student Focus					
Students On Task	4, 310	1.46	.213	--	--
Students Off Task	4, 310	5.06	.001*	.25	PTT < PTNT, NPTT
Students Out of Room	4, 310	5.61	.000*	.28	Comp. > PTT, PTNT, NPTT
Students Waiting	4, 310	2.19	.070	--	--
Teacher Focus					
Students On Task	4, 310	3.55	.008*	.21	Comp. > NPTT
Students Off Task	4, 310	5.34	.000*	.26	PTT < other four groups
Students Out of Room	4, 310	5.65	.000*	.27	Comp. > other four groups
Students Waiting	4, 310	1.37	.246	--	--

*Sig. < .05

Table 10
Percent of Students in SSOS Student Engagement Categories by Grouping

Group	On Task	Off Task	Out of Room	Waiting
PTT	92%	2%	5%	1%
PTNT	84%	8%	3%	5%
NPTT	77%	12%	3%	8%
NPTNT	81%	6%	5%	8%
Comp.	81%	4%	9%	6%

Classroom snapshot: Groups and activities. This section of the COF focused on the teacher, aide, or student groups, as well as the activity taking place (interaction, working alone, management, or social/uninvolved).

One-way ANOVAs were generated to determine whether statistically significant differences existed within these variables (interactive, work alone, management, or social for teacher, aide, and student groups by either target student or teacher focus) by grouping. Appropriate post hoc procedures (Dunnett's *C*, Tukey) were applied to determine which groups were significantly different from one another for the three ANOVAs that did identify significant differences by grouping. For target-student focus,

the student group involved socially resulted in a significant difference by group: the PTT group had significantly fewer students involved in social activities or engaged in nonacademic activities than the NPTT group. For teacher focus, the PTNT group had significantly fewer students working alone than the comparison group, and the PTT group had significantly fewer students involved in social activities or engaged in nonacademic activities than any other group. All three effect sizes were medium (.33, .30, and .33, respectively), indicating there was some practical significance. See Table 11 for a summary of these results.

Table 11
One-Way ANOVA Results for Numbers of Students in SSOS Groups and Activities by Target Student or Teacher Focus

Groups and Activities Category	*df*	*F*	*p*	Effect size	Difference
Target Student Focus					
Students Involved Socially or Uninvolved Academically	4, 133	3.76	.006*	.33	PTT < NPTT
Teacher Focus					
Students Working Alone	4, 134	3.04	.020*	.30	PTNT < Comp.
Students Involved Socially or Uninvolved Academically	4, 140	4.01	.004*	.33	PTT < other four groups

*Sig. < .05

Finally, the percent of students engaged in each type of activity (interaction, working alone, management, or social/uninvolved) was calculated by the five groupings (PTT, PTNT, NPTT, NPTNT, Comp.). In this instance, interactive instruction was defined as a measure of the percentage of students interacting with a teacher, aide, or other students on academic subjects. According to Stallings (1980), interactive instruction rates above 50% have been associated with highly effective classroom instruction. For each of the five groups, this interactive instruction percentage was above 50%, and the PTT group had 75% interactive instruction. See Table 12 for a summary of activity percentages by grouping.

Table 12
Percent of Students in SSOS Groups and Activities by Grouping

Group	Interaction	Work Alone	Management	Social/Uninvolved
PTT	75%	14%	8%	3%
PTNT	58%	24%	6%	12%
NPTT	52%	20%	10%	17%
NPTNT	61%	23%	8%	8%
Comp.	58%	27%	8%	7%

Ongoing activities. This section of the SSOS indicates which of 27 discrete activities was being observed during each 8-minute block (up to seven blocks, four

teacher focused and three student focused) and how many minutes (in whole minute increments from 1 to 8) were spent engaged in a particular activity. For both target student and teacher focus, the most frequent activity was independent inquiry by students; the least frequent was discipline for the target student focus and student-initiated questions.

The 27 discrete activities were grouped into four main categories of teacher led, management/organization, student led, and off task. Table 13 provides the classification of each item into one of these four categories.

Table 14 displays the descriptive statistics for numbers of minutes spent In SSOS main activity categories by grouping. On a global level, the average number of minutes spent on teacher-led activities when the focus was the target student was 10.93 with a standard deviation (*SD*) of 5.65; for the teacher focus, the average was 10.37 (*SD* = 5.78). The average number of minutes spent on target student-focused management/ organization activities was 2.71 (*SD* = 1.97); for the teacher focus, the average was 5.17 (*SD* = 3.54). For target student-focused student-led activities, the average number of minutes was 8.63 (*SD* = 5.28); for the teacher focus, the average was 6.84 (*SD* = 4.63). For target student-focused off task activities, the average was 5.54 (*SD* = 3.99); for the teacher focus, the average was 6.70 (*SD* = 4.61). In general, the teacher-led activities were most prevalent for both target student and teacher focus across all five groups.

One-way ANOVAs were generated to determine whether statistically significant differences existed within these eight variables (teacher led, management/organization, student led, and off task by either target student or teacher focus) by grouping (PTT, PTNT, NPTT, NPTNT, Comp.). Three of the eight ANOVAs did identify significant differences by grouping, and an appropriate post hoc procedure (Dunnett's *C*) was selected to identify which groups were statistically significantly different from one another. For target student-focus, the two categories of student led and off task resulted in statistically significant differences by group: for student led, PTNT > PTT; for off task, NPTT > PTT and Comp. For teacher focus, one category of teacher led was statistically significant by group: NPTT < PTT and Comp. With two small and one moderate effect sizes in the low .20's and low .30's, these findings indicate that there was practical significance. See Table 15 for a summary of results.

QAIT. The QAIT assessment of classroom instrument measured four features of the classroom: quality of instruction, appropriate level of instruction, incentives for learning, and use of time. Forty specific features were rated on a 1 to 5 Likert-type scale (1 = *unlike this class* to 5 = *like this class*).

Table 16 provides the QAIT categories' data by grouping. The PTT and Comp. groups scored highest on the teacher using effective management (means of 4.58 and 4.51, *SD*s of 0.75 and 1.06), the PTNT and NPTT groups on necessary time being allocated for instruction (means of 3.88, *SD*s of 1.35 and 1.08), and the NPTNT group on teachers using an appropriate pace to cover content (mean of 3.89, *SD*s of 1.27). Four of the groups (PTT, NPTT, NPTNT, and Comp.) scored lowest on the teacher using

academic incentives such as small groups with individual incentives (means of 1.06, 1.28, 1.06, and 1.13, respectively, with *SD*s of 0.24, 0.78, 0.25, and 0.58, respectively). The PTNT group scored lowest on the teacher using extrinsic behavioral incentives such as tokens and rewards for improvement (mean of 1.29, *SD* of 0.80).

Table 13
Classroom Observation Individual Activities by Main Categories

Main Categories	Individual Activities
Teacher-Led	A. Teacher presentation of content B. Recitation or discussion C. Directions for assignments D. Small-group instruction E. Tests F. Checking M. Praising class
Management/Organization	G. Procedural or behavioral presentation H. Administrative routines I. Transitions N. Monitoring
Student-Led	O. Individual seatwork P. Individual seatwork at computer Q. Pairs or group seatwork R. Pairs or group seatwork at computer S. Sustained writing or composition T. Sustained reading U. Hands-on learning V. Independent inquiry or research W. Student-initiated questions
Off Task	J. Teacher nonacademic activity K. Waiting time L. Discipline X. Student nonacademic activity Y. Not occupied Z. Off task ZZ. Out of room

Table 14
Descriptive Statistics for Numbers of Minutes in SSOS Main Activity Categories by Target Student or Teacher Focus and by Grouping

Group	Statistic	Teacher Led	Manage./ Organize.	Student Led	Off Task
Target Student Focus					
Pilot Teachers Teaching CRU*	*N* observed	32	16	26	19
	Mean	12.72	3.31	6.19	4.26
	SD	5.30	1.96	3.42	2.83
Pilot Teachers Not Teaching CRU	*N* observed	34	19	34	28
	Mean	9.94	2.74	10.62	6.07
	SD	6.76	1.70	6.29	4.26
Non-Pilot Teachers Teaching CRU	*N* observed	29	11	25	29
	Mean	9.59	3.18	8.60	8.31
	SD	4.87	1.60	4.67	5.54
Non-Pilot Teachers Not Teaching CRU	*N* observed	43	28	35	36
	Mean	10.35	2.39	9.03	6.33
	SD	5.51	1.77	5.24	4.14
Comparison Teachers	*N* observed	141	80	112	110
	Mean	11.22	2.64	8.46	4.63
	SD	5.56	2.15	5.27	3.11
Teacher Focus					
Pilot Teachers Teaching CRU*	*N* observed	31	17	25	16
	Mean	11.65	6.65	6.00	5.19
	SD	4.90	5.23	4.12	4.37
Pilot Teachers Not Teaching CRU	*N* observed	34	20	34	22
	Mean	9.32	4.40	8.47	7.14
	SD	5.66	2.44	5.47	4.54
Non-Pilot Teachers Teaching CRU	*N* observed	29	22	22	26
	Mean	7.62	4.23	6.00	7.88
	SD	4.40	2.47	3.46	3.99
Non-Pilot Teachers Not Teaching CRU	*N* observed	37	26	29	34
	Mean	9.89	6.12	7.10	7.97
	SD	5.30	4.13	4.60	4.91
Comparison Teachers	*N* observed	130	102	100	91
	Mean	11.09	5.03	6.60	6.04
	SD	6.20	3.35	4.61	4.61

*CRU = culturally responsive unit

Table 15
One-Way ANOVA Results for Numbers of Minutes in SSOS Main Activity Categories by Target Student or Teacher Focus

Category	*df*	*F*	*p*	Effect size	Difference
Target Student Focus					
Teacher Led	4, 274	1.70	.151	--	--
Management/Organization	4, 149	0.73	.570	--	--
Student Led	4, 227	2.75	.029*	.22	PTNT > PTT
Off Task	4, 217	6.48	.000*	.34	NPTT > PTT, Comp.
Teacher Focus					
Teacher Led	4, 256	2.96	.020*	.21	NPTT < PTT, Comp.
Management/Organization	4, 182	1.91	.111	--	--
Student Led	4, 205	1.55	.190	--	--
Off Task	4, 184	2.06	.088	--	--

*Sig. < .05

The 40 items were grouped into the four main features of the QAIT: quality of instruction, appropriate level of instruction, incentives for learning, and use of time. Table 16 provides descriptive statistical information for each of the four subscales by groups (PTT, PTNT, NPTT, NPTNT, Comp.).

Table 16
Descriptive Statistics for SSOS QAIT Categories by Grouping

Group	Statistic	Quality of Instruction	Appropriate Level of Instruction	Incentives for Learning	Use of Time
Pilot Teachers Teaching CRU*	*N* observed	33	33	33	33
	Mean	4.09	2.18	2.95	4.45
	SD	0.64	0.69	0.62	0.67
Pilot Teachers Not Teaching CRU	*N* observed	42	42	42	42
	Mean	3.47	2.17	2.56	3.74
	SD	1.20	0.86	0.88	1.41
Non-Pilot Teachers Teaching CRU	*N* observed	36	36	36	36
	Mean	3.28	2.10	2.53	3.69
	SD	0.97	0.90	0.81	0.95
Non-Pilot Teachers Not Teaching CRU	*N* observed	46	47	47	47
	Mean	3.26	1.54	2.18	3.77
	SD	1.11	0.66	0.79	1.19
Comparison Teachers	*N* observed	151	153	153	151
	Mean	3.70	1.97	2.66	4.19
	SD	0.96	0.64	0.71	1.07

*CRU = culturally responsive unit

One-way ANOVAs were generated to determine whether statistically significant differences existed within these subscale variables (quality of instruction, appropriate level of instruction, incentives for learning, and use of time) by grouping (PTT, PTNT, NPTT, NPTNT, Comp.). All four ANOVAs did identify significant differences by grouping, and an appropriate post hoc procedure (Dunnett's *C*) was selected to identify which groups were statistically significantly different from one another. With moderate effect sizes all at .25 or above, the statistical significance was accompanied by practical significance. See Table 17 for a summary of these results.

Table 17
One-Way ANOVA Results for SSOS QAIT Categories

Category	*df*	*F*	*p*	Effect size	Difference
Quality of Instruction	4, 303	4.87	.001*	.25	PTT > all other groups
Appropriate Level of Instruction	4, 306	6.12	.000*	.28	NPTNT < all other groups
Incentives for Learning	4, 306	5.86	.000*	.28	NPTNT < PTT, Comp.
Use of Time	4, 304	5.19	.001*	.25	PTT > PTNT, NPTT, NPTNT; NPTT < Comp.

*Sig. < .05

CERC. The Classroom Environment and Resources Checklist (CERC) assesses the presence or absence of indicators of good classroom environments, as well as the visibility and use of a variety of resources. A CERC form was completed at the end of each of the 315 classroom observations. Table 18 presents the percentages for the presence of 14 environmental indicators by the five groups (PTT, PTNT, NPTT, NPTNT, Comp.). For the PTT group, adequate lighting and cheerful and inviting classroom were the most frequently seen environmental indicators; least seen was distinct activity centers. For the PTNT group, adequate lighting and posted classroom rules were most frequently seen; least seen was student-controlled classroom discourse. For both the NPTT and NPTNT groups, the most frequently seen indicators were adequate lighting and comfortable ventilation/temperature; least seen was student-controlled classroom discourse. For the Comp. group, adequate lighting and posted classroom rules were most frequently seen; least seen was culturally mediated instruction.

Table 19 presents the percentages depicting use of 18 resources by the five groups (PTT, PTNT, NPTT, NPTNT, Comp.). For all five groups, the most often used resource was the classroom chalkboard. For the PTT group, least often used was a science/lab table; for the PTNT group, a map and/or globe; for the NPTT group, reference materials and a science/lab table; for the NPTNT group, games and/or puzzles and student-used equipment; and for the Comp. group, a science/lab table.

Table 18
Numbers and Percentages for Presence of SSOS CERC Environmental Indicators by Grouping

Items	Statistic	Grouping				
		PTT	PTNT	NPTT	NPTNT	Comp.
Culturally-mediated instruction	*N* Percent	10 30%	9 21%	10 28%	6 13%	2 1%
Student-controlled classroom discourse	*N* Percent	8 24%	7 16%	5 14%	3 6%	15 10%
Use of multi-racial materials	*N* Percent	17 52%	11 26%	16 44%	24 51%	30 19%
Use of non-sexist materials	*N* Percent	21 64%	21 49%	29 81%	28 60%	75 48%
Posted classroom rules	*N* Percent	30 91%	35 81%	30 83%	41 87%	148 95%
Posted assignments	*N* Percent	16 48%	29 67%	22 61%	23 49%	112 72%
Cheerful and inviting classroom	*N* Percent	32 97%	26 60%	26 72%	40 85%	137 88%
Distinct activity centers	*N* Percent	3 9%	11 26%	8 22%	26 55%	65 42%
Adequate lighting	*N* Percent	33 100%	41 95%	36 100%	43 92%	151 97%
Comfortable ventilation/ temperature	*N* Percent	31 94%	31 72%	31 86%	44 94%	135 86%
Student work displayed	*N* Percent	26 79%	22 51%	9 25%	20 43%	94 60%
No distracting internal noises/ interruptions	*N* Percent	25 76%	20 46%	21 58%	31 66%	105 67%
No distracting external noises/ interruptions	*N* Percent	23 70%	26 60%	26 72%	26 55%	98 63%
Open, risk-free environment	*N* Percent	29 88%	30 70%	31 86%	42 89%	139 89%

Table 19
Numbers and Percentages for Use of SSOS CERC Resources by Grouping

Items	Statistic	Grouping				
		PTT	PTNT	NPTT	NPTNT	Comp.
Textbooks	*N*	13	20	19	26	102
	Percent	39%	46%	53%	55%	65%
Workbooks/activity books	*N*	2	2	2	9	33
	Percent	6%	5%	6%	19%	21%
Worksheets/activity sheets	*N*	15	18	20	27	96
	Percent	46%	42%	56%	57%	62%
Journals/learning logs	*N*	13	8	6	7	28
	Percent	39%	19%	17%	15%	18%
Classroom library	*N*	6	3	4	5	25
	Percent	18%	7%	11%	11%	16%
Reference materials	*N*	6	3	0	4	14
	Percent	18%	7%	0%	8%	9%
Map and/or globe	*N*	4	0	3	1	8
	Percent	12%	0%	8%	2%	5%
Games and/or puzzles	*N*	2	3	1	0	9
	Percent	6%	7%	3%	0%	6%
Instructional aids/props	*N*	8	5	5	8	27
	Percent	24%	12%	14%	17%	17%
Science/lab table(s)	*N*	0	7	0	2	0
	Percent	0%	16%	0%	4%	0%
Classroom chalkboard	*N*	25	30	22	29	121
	Percent	76%	70%	61%	62%	78%
Student-used equipment	*N*	3	9	2	0	6
	Percent	9%	21%	6%	0%	4%
Overhead projector	*N*	7	9	4	10	49
	Percent	21%	21%	11%	21%	31%
Television	*N*	1	2	4	5	8
	Percent	3%	5%	11%	11%	5%
Computer	*N*	6	8	7	7	35
	Percent	18%	19%	19%	15%	22%
Student manipulatives/ hands-on materials	*N*	5	7	3	9	23
	Percent	15%	16%	8%	19%	15%
Audio resources	*N*	3	2	7	2	4
	Percent	9%	5%	19%	4%	3%
Video resources	*N*	1	4	4	4	3
	Percent	3%	9%	11%	8%	2%

Adherence index. This score reflects observed teachers' adherence to the principles of culturally responsive instruction. Twenty-two items from the QAIT and CERC that most closely aligned to one of the nine specific components of the culturally responsive units were used in this analysis to form *z* scores for each component, which were then transformed to a standardized *Z* score variable (refer back to Table 4 to see which items make up each component score).

A one-way ANOVA was generated to determine whether statistically significant differences existed among the five groups for this adherence variable; the ANOVA was significant ($F(4, 310) = 8.21$, $p < .05$). An appropriate post hoc procedure (Dunnett's *C*) was selected to identify which groups were significantly different from one another. The pilot team teachers who were teaching one of the units had a significantly higher adherence score than the NPTT, NPTNT, and Comp. group, but not significantly higher than the PTNT group. Further, the comparison group of teachers also had a significantly higher score than the nonpilot team teachers who did not teach such a unit. The medium effect size of .32 indicates some practical significance. Figure 2 provides a visual depiction of the *Z* scores for each of the five groups. This figure also shows that the pilot team teachers who were teaching one of the culturally responsive units had a much higher adherence index score than any of the other four groups.

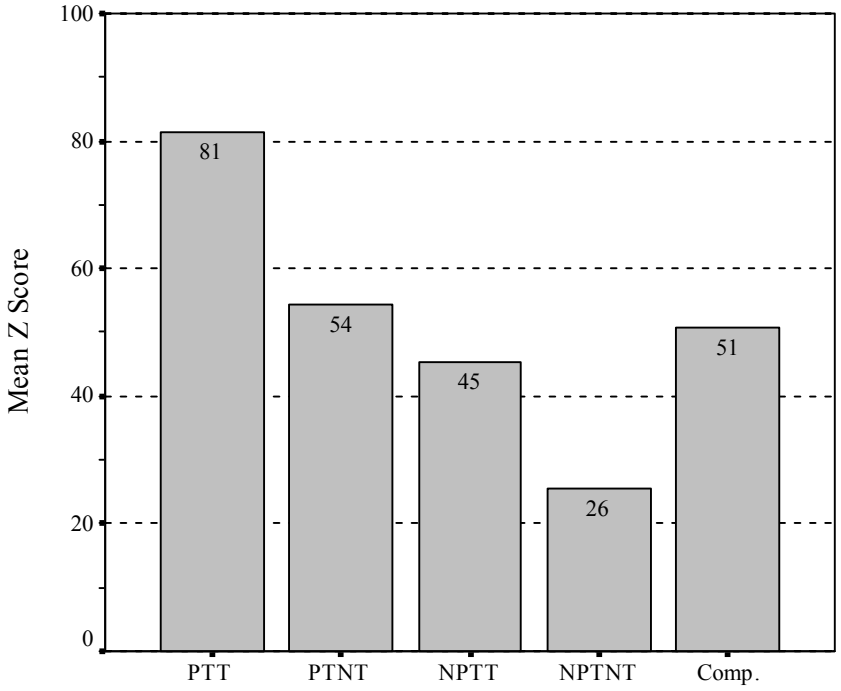

Figure 2
SSOS Culturally Responsive Instruction Adherence Index Scores by Grouping

AEL Measure of Academic Supportiveness and Climate

The four subscales of the AEL MASC contain 19 (Student Belonging), 6 (Family Expectations), 7 (Student Academic Efficacy), and 7 (Family/School/Student Involvement) items, which respondents rated using a scale ranging from 1 to 5 (*not at all true* to *always true*). A mean score ranging from 1 to 5 was calculated for each subscale. Table 20 provides descriptive statistical summaries for the four subscales for the pilot and comparison schools at the full group level and at the building-level.

Pilot schools generally had higher subscale mean scores across three of the four scales than did the comparison schools. When looking at the data for the full group, statistically significant differences were found on two of the four subscales (Student Belonging and Family Expectations). The difference favored the comparison schools on the former subscale and the pilot schools on the latter. Small effect sizes were found.

Table 21 provides descriptive statistical summaries for the four subscales across two administrations of the AEL MASC to students in the pilot school. The spring 2004 administration showed higher mean scores across the four subscales than did the spring 2003 administration. Independent *t* tests were computed comparing both administrations on the four subscales. Statistically significant differences were found on three of the four subscales: Student Belonging, Family Expectations, and Student Academic Efficacy.

Student Achievement Assessment

AEL researchers collected WESTEST data for 249 students observed at pilot schools and 362 students observed at comparison schools. Of the pilot school students, 96 were enrolled at an elementary school, 148 were enrolled at a middle school, and 7 were high school students. Among comparison school students, 205 were elementary-level students, 121 were middle school students, and 36 were high school students.

For mathematics performance, at every grade level except Grade 4, comparison schools had a higher percentage of students achieving or exceeding mastery level than did pilot schools. In Grade 4, data were available for only 6 pilot school students, all of whom achieved at least the mastery level on the WESTEST for mathematics. Slightly less than half of the comparison students in the 4th grade achieved or exceeded mastery in mathematics; in all other grade levels more than half of the students achieved mastery. Only in Grades 5 and 6 did at least half of the pilot school students meet standards for mastery in mathematics.

For reading and language arts, more than half of comparison school students at all grade levels met or exceeded the standards for mastery level performance, and more than 90% of 10th-grade comparison school students were at or above mastery. More than half of pilot school students meet or exceed mastery level performance for all grade levels except Grade 3. Comparison schools continued to have a higher rate of students achieving or exceeding mastery (except for 4th grade in which only 6 pilot school students were observed). However, pilot school students in Grade 6 achieved mastery at a rate comparable to their comparison school peers (73% and 75%).

As with performance in mathematics and reading/language arts, all six pilot school fourth-grade students achieved or exceeded mastery level in science, and slightly less than half of pilot school third-grade students met standards for mastery. More than half of students in the remaining grade levels (grades 5, 6, 7, and 10) at the pilot schools met or exceeded the mastery level. At all grade levels in the comparison schools, more than half of the students were at or above mastery for science. Pilot school students in Grades 7 and 10 achieved mastery at rates comparable to their counterparts in

Table 20
AEL MASC Subscale Descriptive Statistics and Differences by Full Group and Building Level

Subscale Name	Level	Pilot Schools			Comparison Schools			*df*	*t*	*p*	Dif.	*d*
		N	Mean	*SD*	*N*	Mean	*SD*					
Student Belonging	Elementary	201	3.94	0.71	363	4.03	0.68	395.47	1.49	ns	-0.09	0.13
	Middle	460	3.53	0.73	313	3.43	0.87	587.83	1.67	ns	0.10	0.12
	High	691	3.44	0.72	758	3.55	0.75	1447	2.93	.003	-0.11	0.15
	Full Group	1352	3.54	0.74	1434	3.64	0.80	2783.35	3.52	.000	-0.10	0.13
Family Expectations	Elementary	200	4.70	0.62	363	4.68	0.70	561	0.33	ns	0.02	0.03
	Middle	456	4.58	0.63	313	4.57	0.68	767	0.14	ns	0.01	0.01
	High	691	4.60	0.65	756	4.46	0.76	1438.60	3.82	.000	0.14	0.20
	Full Group	1347	4.61	0.64	1432	4.54	0.73	2761.12	2.61	.009	0.07	0.10
Student Academic Efficacy	Elementary	201	4.09	0.76	363	4.02	0.74	562	1.09	ns	0.07	0.10
	Middle	460	3.83	0.76	313	3.68	0.82	771	2.66	.008	0.15	0.19
	High	691	3.75	0.70	758	3.79	0.76	1446.56	0.93	ns	-0.04	0.05
	Full Group	1352	3.83	0.74	1434	3.82	0.78	2784	0.25	ns	0.01	0.01
Family/School/Student Involvement	Elementary	201	4.38	0.66	363	4.30	0.68	562	1.41	ns	0.08	0.12
	Middle	460	3.95	0.75	313	3.92	0.80	771	0.51	ns	0.03	0.04
	High	691	3.58	0.81	758	3.51	0.85	1447	1.52	ns	0.07	0.08
	Full Group	1352	3.82	0.82	1434	3.80	0.86	2783.71	0.73	ns	0.02	0.03

Table 21
AEL MASC Subscale Descriptive Statistics and Differences for Pilot Schools by Year of Administration

Subscale Name	Spring 2003			Spring 2004			*df*	*t*	*p*	Dif.	*d*
	N	**Mean**	***SD***	***N***	**Mean**	***SD***					
Student Belonging	1178	3.48	0.78	1352	3.54	0.74	2435.83	2.16	.031	0.07	0.09
Family Expectations	1169	4.46	0.73	1347	4.61	0.64	2340.24	5.45	.000	0.15	0.22
Student Academic Efficacy	1175	3.72	0.73	1352	3.83	0.74	2525	3.61	.000	0.11	0.14
Family/School/Student Involvement	1176	3.80	0.81	1352	3.82	0.82	2526	0.69	ns	0.02	0.03

comparison schools (67% and 86% to 68% and 89%, respectively). Pilot school students in Grade 6 met or exceeded mastery at a slightly higher rate than their peers at comparison schools (80% to 78%, respectively).

Student achievement comparisons. Using the WESTEST scale scores, AEL researchers compared performance in each content area by grade level for pilot and comparison schools. Table 22 provides descriptive statistical summaries for those analyses. In general, comparison schools had higher mean scale scores across the grades than did the pilot schools.

Independent *t* tests were computed comparing pilot schools and comparison schools at each grade level for the three content areas. Statistically significant differences were found for 3rd-grade students in two of the content areas (mathematics and science), with moderate and large effect sizes. At the 4th-grade level, a statistically significant difference was found for one content area (mathematics), which had a large effect size. At the 5th-grade level, a statistically significant difference with a small effect size was found for one content area (reading/language arts). A statistically significant difference with a small effect size was also found for one content area (mathematics) for 6th-grade students. Statistically significant differences with small to moderate effect sizes were found for all three content areas for 7th-grade students. Statistically significant differences with large effect sizes were found for 10th-grade students in two of the content areas (reading/language arts and science).

Interview Design Process

The Interview Design process conducted at the spring 2004 training workshop included five questions, and the results presented in this section are summarized by question.

Question 1: *Explain two or three things you have learned about teaching African American students that you feel every teacher needs to understand to be effective, and explain why you feel those things are important.*

Respondents most frequently mentioned respecting students' culture and cultural differences as an important lesson learned for teaching African American students. Most responses in this category noted the importance of cultural sensitivity and understanding and incorporating students' culture into instruction. Responses to this question mentioned a host of other lessons respondents felt were important, including the importance of building positive relationships between teachers and students, the need to address multiple learning styles, the necessity of having high expectations for all students, and the need for teachers to have positive attitudes.

Table 22
WESTEST, Student Achievement Data Descriptive Statistics and Differences by Grade Level

Content Area	Grade Level	Pilot Schools			Comparison Schools			*df*	*t*	*p*	Dif.	*d*
		N	Mean	*SD*	*N*	Mean	*SD*					
Mathematics	Third	19	586.79	37.22	77	609.94	30.74	94	2.82	.006	-23.15	0.68
	Fourth	6	657.83	22.62	51	624.65	35.40	55	2.23	.030	33.19	1.12
	Fifth	71	647.56	32.20	77	653.60	44.15	146	0.94	ns	-6.03	0.16
	Sixth	64	636.47	121.34	64	673.73	28.54	69.95	2.39	.019	-37.27	0.42
	Seventh	84	656.74	36.47	57	676.88	29.14	139	3.48	.001	-20.14	0.61
	Tenth	7	705.00	16.03	36	725.25	27.61	41	1.87	ns	-20.25	0.90
Reading/ Language Arts	Third	19	594.68	61.26	77	618.12	43.80	22.74	1.57	ns	-23.43	0.44
	Fourth	6	656.00	20.93	51	610.61	97.67	55	1.13	ns	45.39	0.64
	Fifth	71	638.07	40.32	77	651.86	35.75	146	2.21	.029	-13.79	0.36
	Sixth	64	639.33	119.73	64	664.02	32.38	126	1.59	ns	-24.69	0.28
	Seventh	84	636.19	89.304	57	661.19	36.43	118.17	2.30	.023	-25.00	0.37
	Tenth	7	676.14	21.39	36	704.58	20.25	41	3.37	.002	-28.44	1.37
Science	Third	19	596.58	41.87	77	625.30	28.94	94	3.52	.001	-28.72	0.80
	Fourth	6	648.50	26.22	51	603.29	127.53	55	0.86	ns	45.21	0.49
	Fifth	71	635.80	38.79	77	644.61	33.41	146	1.48	ns	-8.81	0.24
	Sixth	64	639.94	145.94	64	655.45	89.72	126	0.73	ns	-15.52	0.13
	Seventh	84	643.52	129.10	57	676.32	25.12	92.09	2.27	.026	-32.79	0.35
	Tenth	7	684.86	30.97	36	709.14	21.68	41	2.53	.015	-24.28	0.91

Question 2: *Name one or two decisions, made outside the classroom, about how students experience and progress through school (e.g., course offerings, discipline procedures, scheduling methods) that you feel either hinder or help African American students reach their potential. Explain why you feel they hinder or help.*

Respondents offered a variety of responses to this item, and most responses focused on factors that hinder African American students' ability to reach their full potential; only a very few responses made mention of factors that help. Such distribution of responses might indicate that, in the opinions of the Interview Design participants, very few of the decisions affecting African American students are helping those students reach their full potential or achieve at higher levels. Respondents specifically mentioned as areas of concern: decisions relating to discipline issues, classroom placement, extracurricular activities (e.g., sports, clubs), opportunities for students to enroll in higher level classes, schoolwide scheduling procedures and issues, weak support systems for African American students, and teacher attitudes and beliefs

Question 3: *Think of a time in the past year when you witnessed an African American student or students working at a level beyond what had been typical. Describe the circumstances. To what do you attribute the student(s)' higher level of work?*

For the most part, participants in the Interview Design process attributed higher levels of student work to student factors (e.g., focus, excitement about a particular topic or project, motivation to excel, goal setting, practice or preparation) and teacher factors (e.g., teacher involvement, instruction, support, encouragement, high expectations, challenging students to work at higher levels).

Question 4: *What do you believe are the two or three most important factors that contribute to the achievement gap, and why?*

All responses indicated that participants acknowledged an achievement gap between African American students and their peers of other ethnicities. Participants believed that some parents might not be interested or involved in their children's education, might be unable to help because of work responsibilities, or might lack the resources (e.g., reading material) to encourage high educational attainment. Some respondents felt that some teachers might be unwilling to break away from currently used ineffective or culturally insensitive teaching practices, perhaps because those teachers are afraid to do so or are unwilling to learn new teaching styles. Other respondents cited lack of appropriate teacher training and materials, some teachers' low expectations for or negative attitudes about African American students, and some teachers' inexperience or unfamiliarity with teaching African American students.

Question 5: *What do you consider to be two or three instructional strategies that hold the most promise for narrowing the achievement gap between African American/low-SES students and other students, and why?*

Participants offered many different responses to this item, indicating that participants did not view one or two strategies as being vastly more promising than all others in terms of narrowing achievement gaps. Some of strategies that were mentioned in response to this question were reading (including reading across the curriculum, sustained silent reading, directed reading, and accelerated reading programs), ensuring that instruction and instructional materials were culturally relevant, employing cooperative learning strategies, and addressing multiple learning needs.

Comparison: 2001 and 2004

Results of focus groups conducted in late 2001 with teachers at the five schools then participating in the Pilot Schools project were compared to the 2004 Interview Design responses. In both 2001 and 2004, participants were asked what factors contributed to the achievement gap. In both years, participants posited that all of the following contributed to the achievement gap: students' home and family situations, family and parent attitudes toward education, lack of family and parent support or involvement, and placing a low priority on education. Respondents in both years also contended that student beliefs and attitudes, teacher attitudes and related factors, and low expectations for African American students also accounted for differences in level of achievement. Cultural factors were mentioned as contributors at both times, but these responses were slightly more pronounced in the 2001 focus groups.

Although participants in 2004 mentioned student beliefs and attitudes as contributing factors, 2001 focus group respondents did so at a slightly higher rate. Additionally, though respondents in some of the 2001 focus groups did mention teacher attitude and teaching style as contributors to the achievement gap, participants in the 2004 Interview Design process seemed to place more weight on this factor. These slight differences suggest a shift in perception of responsibility and accountability over the years from 2001 to 2004.

In 2001, respondents seemed hesitant to acknowledge that there was an achievement gap between African American students and students of other ethnicities. Respondents wanted to attribute any achievement gap to socioeconomic disparities, not ethnicity. Focus group participants responded that they "don't see color" or "don't need to focus on color" or race when instructing their students. Although participants in the 2004 interview design acknowledged socioeconomic status as a factor that contributed to the achievement gap, they seemed more willing to acknowledge that there was an achievement gap between different ethnic groups. Respondents in 2004 also seemed much more willing to admit that students of different ethnicities might require different teaching or instructional strategies.

DISCUSSION AND CONCLUSIONS

The following section presents discussion and conclusions drawn from the findings of the current research project. The conclusions, preceded by bulleted summaries of findings, address the seven research questions posed at the outset of this project.

Question 1. ***What is the effect on student achievement of schools' participation in a pilot research and development project designed to be responsive to the cultural needs of African American and low-socioeconomic-status (SES) students?***

- Overall, pilot and comparison students were comparable in terms of achieving mastery in the subject areas of reading/language arts and science. Comparison students, however, showed greater mastery of math than did the pilot students.

- In general, at each grade level except fourth grade, comparison students outperformed pilot students with few exceptions. Thus, comparison school students observed in this research generally performed better than pilot students in terms of student achievement.

It would be inappropriate to make a claim of gain or loss in students' achievement, as measured on standardized tests, based on the very circumscribed implementation of culturally responsive teaching practices within each school and the teaching of one culturally responsive unit. That we could not examine student achievement data over time was an artifact of the timing of the study rather than a flaw in the research design.

If we had been afforded the opportunity to compare students' standardized test scores across the years, as they were exposed to multiple teachers within their schools using culturally responsive teaching practices and lessons, we expect that we might have been able to show movement in pilot school students' achievement scores, given the findings of improvement in other areas we examined over time. We would also need to ensure continuity of assessment instruments. In the current case, the state adopted a new criterion-referenced achievement test (WESTEST), which was administered to Grades 3 through 8 and Grade 10 during the 2003-04 school year. Prior to that time, the state used the SAT-9, a norm-referenced test, administered to Grades 3 through 11.

We do know that when the culturally responsive units were taught by pilot team teachers, significantly fewer students were coded as off task. We also know that pilot team teachers and students using the culturally responsive units spent fewer minutes per hour engaged in off-task activities when compared to other groups. This finding, in light of the importance of student engagement in learning, is particularly promising. On-task behavior, of course, is associated with higher student achievement and increased levels of learning (Stallings, 1980).

Our data show that teachers who learn about culturally responsive teaching practices and teach standards-based lessons designed to be consistent with culturally

responsive teaching principles are more likely to keep all students, including African American students, on learning tasks during the school day. Therefore, over time, continued participation in this intervention should produce higher student achievement due to less time spent off task.

Question 2. ***In what ways, if at all, does schools' participation in a pilot research and development project to improve the achievement of children, especially African American and low-SES students, affect school staffs' sense of themselves as a high-performing learning community?***

- At the full-group level, results from the 2004 administration of the AEL CSIQ show that comparison schools were significantly more committed to continuous learning and improvement on the subscales having to do with school/family/community connections, shared goals for learning, and effective teaching.

- At the building level, the pilot and comparison high schools were equally committed to continuous school improvement on five of the six subscales (i.e., Learning Culture, School/Family/Community Connections, Shared Goals for Learning, Effective Teaching, and Purposeful Student Assessment).

- On the sixth subscale, Shared Leadership, the pilot high school showed greater commitment to improvement than the comparison school.

- On the subscale Purposeful Student Assessment, pilot and comparison schools at all levels were equally committed.

To understand the results of school staffs' perceptions of themselves as high-performing learning communities, it is important to learn how the three groups of schools in the study were chosen. The four pilot schools—two elementary schools, one middle school, and one high school—were part of the same feeder pattern. Three of the schools, the elementary and middle schools, were chosen by the superintendent to participate because they were low performing by state standards in effect during 2001-2002 and because their student bodies included high percentages of both low-income students and African American students. Of the leadership in those three schools, one was enthusiastic about participating, one was compliant, and one was resistant. The high school volunteered to participate. It had not been designated as a low-performing school, but it did have a significant percentage of African American students.

The four comparison schools were chosen from other Kanawha County schools based on demographics that were as close as possible to the pilot schools. None of the four was designated low-performing. Although the percentages of students receiving free or reduced-price lunch were similar in all eight schools, the comparison schools had lower percentages of African American students and higher test scores than the pilot schools.

The field-tested and validated AEL Continuous School Improvement Questionnaire (AEL CSIQ) assesses elements associated with high-performing learning cultures. It is to be expected, therefore, that schools identified as low-performing would reflect this reality in their scores on the AEL CSIQ and that the comparison schools, which were not low-performing, would show a higher commitment generally to continuous school improvement. The pilot high school, not a low-performing school, was similar in scores with the comparison high school on most of the subscales. So, this is not unexpected as it helps to confirm why the other three pilot schools revealed themselves to be less committed to components of high-performing learning communities. The fact that schools are judged by student progress on state-mandated tests makes it unsurprising that all schools were equally committed to purposeful student assessment.

Question 3. ***In what ways, if at all, does schools' participation in this pilot research and development project to improve the achievement of children, especially African American and low-SES students, affect schools' capacity to undertake improvement initiatives?***

- Compared to the perceptions of pilot school staff, comparison school professional staff reported that their schools had greater capacity for improvement on four of eight subscales of the AEL Measure of School Capacity for Improvement (AEL MSCI): Collective Professional Capacity, Technical Resources, Differentiated Instruction, and Expectations for Student Performance.

- On four of the eight subscales of the AEL MSCI, pilot schools and comparison schools reported that they were equally ready for improvement. The areas in which the two groups shared similar capacity to improve were Peer-Reviewed Practice, Program Coherence, Anti-Discriminatory Teaching, and Responsive Pedagogy.

- At the building level, staff at the pilot high school reported that their school was slightly better prepared for improvement than the comparison high school on six of the eight subscales (Collective Professional Capacity, Peer-Reviewed Practice, Program Coherence, Anti-Discriminatory Teaching, Responsive Pedagogy, and Differentiated Instruction), with two of those subscales (Peer-Reviewed Practice and Program Coherence) indicating significantly greater capacity to improve.

- Over time, the pilot schools have, as a group, increased their capacity for improvement on six of the eight AEL MSCI subscales: Collective Professional Capacity, Peer-Reviewed Practice, Program Coherence, Technical Resources, Anti-discriminatory Teaching, and Responsive Pedagogy.

- Pilot schools were equal to or lower than comparison schools on half of the AEL MSCI subscales on the 2004 administration; however, pilot school faculty

members' perceptions about their schools' capacity to improve changed in a positive direction over the course of the pilot schools' intervention.

These findings show that the pilot schools intervention helped faculty in pilot schools increase their perceptions of their schools' capacity to improve. The increased capacity for improvement is particularly noticeable in the area of anti-discriminatory teaching, which was the focus of this intervention.

These findings are particularly encouraging, given that the majority of faculty at the pilot schools were not directly involved in the project. One possible explanation for this phenomenon is the role pilot school principals played within their respective schools. Pilot school principals were asked to attend all full-day meetings and encouraged and invited to participate in biweekly pilot team meetings within their respective schools. Thus, the intervention sought to immerse the key gatekeeper to a school's faculty in the intervention.

Although we did not systematically collect evidence about the pilot school intervention and activities and information that were shared with the faculties at the respective pilot schools, we do know anecdotally that the principals gave their permission and, in many instances, requested that pilot team teachers share what they were learning with the entire faculty at the pilot schools. Thus, there was an information "spillover" from the pilot school intervention into the general faculty. The fact that anti-discriminatory teaching showed the greatest improvement is particularly encouraging. An increase in anti-discriminatory teaching should, in fact, create a school environment that supports the learning of all students, including African American students and low-SES students.

Question 4. ***In what ways, if at all, does schools' participation in a pilot research and development project to improve the achievement of children, especially African American and low-SES students, alter students' perceptions of their schools' and families' support for students' academic endeavors and school climate conducive to learning?***

- In terms of students' perceptions of their schools' academic supportiveness and climate, the two groups were relatively the same for two subscales: Student Academic Efficacy and Family/School/Student Involvement.

- Comparison students reported having a greater sense of belonging in their schools, and pilot students reported that they felt greater expectations from their families.

- Over time, pilot students' perceptions of their schools' and families' support for academic endeavors and climate for learning improved significantly in areas related to their perceptions of belonging, their own ability to do well academically, and their families' expectations of them.

- Pilot school students were equal to or greater than comparison school students in their perceptions of academic efficacy, family/school/student involvement, and family expectations.

- Students' perceptions and experiences indicate essentially no differences between pilot and comparison schools. However, faculty members seem to be differentiated in their perceptions of their schools' commitment to continuous learning and readiness to improve in areas related to the pilot schools' intervention.

The pilot school intervention did, in fact, influence students' perceptions in a positive manner. Because those students directly affected by teachers who received the full intervention treatment are a relatively small portion of the full student body, particularly in the middle and high schools, it is remarkable that the three-year project appears to have produced improvements in overall student perceptions in areas directly addressed by the project. The finding that students improved in their perceptions of belonging and their own ability to do well academically may be explained by overall teacher growth in readiness to improve in the areas of anti-discriminatory teaching and responsive pedagogy. Also, because three of the four pilot schools were chosen because they were low performing, and their low performance was reflected in differences between pilot and comparison school teacher perceptions on the AEL CSIQ, it is notable that at the project's end, there were essentially no differences between pilot and comparison school students' perceptions and experiences on the AEL MASC. Evidence supports the conclusion that the intervention was successful in improving students' perceptions of their schools' and families' support for academic endeavors and climate for learning.

An additional factor that may account for some of the change in students' perceptions is the leadership role that principals and pilot school teachers played in their respective schools. As pilot team members, including the principal, became aware of factors that influence students' perceptions of belonging and academic efficacy, they intentionally undertook efforts at the building level to address students' needs.

For example, at one pilot school, the principal reviewed enrollments for advanced math. The principal discovered that there were academically qualified students, African American and White, who for other reasons were not allowed to take the class. The principal then took steps to remedy the situation. Consequently, more students were able to take advantage of more challenging curricula. Efforts such as these may increase students' sense of being able to achieve at higher levels.

In another instance, one pilot school teacher and the principal implemented a "student counselor" program using older students as peer counselors for incoming students, particularly those students they felt were socially isolated. This effort sought to increase students' sense of belonging. It is encouraging that, as a result of their school's involvement in the pilot school intervention, students' perceptions of their schools' supportiveness did improve significantly.

Question 5. ***To what degree does a teacher's participation in the full or partial treatment groups affect the format, substance, and quality of his or her instruction?***

- Pilot team teachers who taught a culturally responsive unit (CRU) demonstrated a higher quality of instruction than all other groups as measured by the Quality of Instruction, Appropriate Level of Instruction, Incentive, and Use of Time (QAIT) instrument.

- Pilot team teachers who taught a CRU had the most positive classroom learning environments as measured by the Classroom Environment and Resources Checklist (CERC) instrument.

- Pilot team teachers had greater adherence to the principles of culturally responsive instruction than non-pilot team teachers and comparison teachers as measured by a specially constructed adherence scale composed of items from the QAIT and CERC instruments.

The findings of this study indicate that, as expected, the teachers who were members of the pilot team had a greater adherence to culturally responsive teaching principles than did non-pilot team teachers. An additional finding showing that pilot team teachers who taught the unit had greater adherence to culturally responsive teaching principles than did pilot team teachers not teaching the unit was unexpected. Because all groups of teachers were observed before, during, and after the unit, it appears that the experience of teaching the unit seemed to reinforce and strengthen a pilot teacher's ability to teach in a culturally responsive manner. Our findings lead us to conclude that providing teachers with information about culturally responsive teaching was not sufficient to produce the greatest adherence or change in instructional practice. However, providing the information and asking teachers to put into practice the principles of culturally responsive teaching did produce the greatest adherence.

In this study, *format* is defined as the degree to which teachers' lessons adhered to the principles of culturally responsive instruction. S*ubstance* in this study includes instructional materials and methods used to create a positive learning environment. *Quality of instruction* in this study refers both to the use of principles of culturally responsive instruction and the use of best instructional practices, such as communicating high expectations, using appropriate pace, and relating topics to students' lives. The format, substance, and quality of instruction in a teacher's class are more likely to conform to the principles of culturally responsive teaching if that teacher has been required to use an exemplar unit that demonstrates principles of culturally responsive teaching. This finding also supports the hypothesis on which the intervention was based: providing teachers with materials and concepts that have immediate applicability to their teaching, in addition to providing ongoing professional development, is a most effective way to increase professional growth.

In terms of instructional format, pilot team teachers teaching a culturally responsive unit demonstrated better format than other groups of teachers. Observation data showed that these teachers had significantly higher adherence to the principles than any other treatment group. Also notable was the fact that teachers in the pilot schools who were not on the team and did not receive the CRU had the lowest adherence score of the four treatment groups and the comparison. This finding is powerful in that regarding the principles of culturally responsive instruction, NPTNT teachers have the most room to improve—even more so than comparison teachers. The fact that the adherence rate for the NPTNT group was the lowest of all treatment groups is not unexpected because this group had little or no opportunity, in some schools, to learn about the principles of culturally responsive instruction and no experience with teaching a CRU.

With regard to substance of instruction, overall, PTT group teachers had the most positive classroom learning environments, especially in their use of culturally mediated instructional activities, student-controlled discourse, and multiracial materials. These three components reflect the principles of culturally responsive instruction; thus, the intervention increased teachers' ability to use learning activities and materials that are culturally responsive. The AEL MSCI scores of pilot school teachers further validate this finding. Over the period of this intervention, the pilot school teachers' perceptions of their readiness to use anti-discriminatory teaching and responsive pedagogy increased significantly.

Furthermore, PTT group teachers were observed to have markedly greater use of journals/learning logs, instructional aids/props, and reference materials and markedly lower use of textbooks, workbooks, and worksheets than other groups of teachers. This is in line with the pilot schools intervention as planned. Use of journals, instructional aids, and reference materials was meant to be part of the CRU, and use of textbooks, workbooks, and worksheets was not meant to be part of the CRU. This is additional evidence to support our conclusion that teachers' participation in the full treatment increased the quality of their instruction.

Also significant about the systematic classroom observation findings related to substance of instruction was that the NPTNT group (who had not participated on the team or taught a CRU) had the least positive classroom learning environments in terms of appropriate levels of instruction and incentives for learning. This is not surprising, considering the fact that the pilot team teachers had attended workshops and bimonthly team meetings that provided them with strategies for creating positive learning environments. Moreover, these PTT teachers had designed lessons during the year that incorporated these strategies, taught them, and participated in reflective conversations with other team members and the facilitator about the effects of these strategies on student engagement and performance. Finally, they taught a culturally responsive unit that contained incentives for learning, a developmentally appropriate time frame and materials, and highly engaging learning activities. Here, again, we have further evidence to support our conclusion that teachers' participation in this full treatment improved the quality of their classroom instruction.

We hypothesized that students will learn more when they interact with materials that are at the appropriate instructional level, over enough time, with enough inherent or external incentives, and presented through high-quality instruction. Overall, the pilot team teachers who taught the CRU demonstrated a significantly higher quality of instruction than all other groups. We conclude that this is attributable to the culturally responsive teaching project, given the fact that the intervention included providing teachers with information about and specific examples of noteworthy instructional and classroom management practices, especially in the CRU they taught. During the monthly team meetings, for example, the facilitator provided specific instructional strategies to illustrate a particular principle of culturally responsive instruction. Pilot team teachers then selected a strategy to implement in a lesson during the following two weeks. At the next team meeting, teachers shared their lessons and discussed their experiences with using the instructional strategies and their students' responses to them. Through classroom application of the strategies and reflection on their effects on student learning, teachers were able to refine their skills in using culturally responsive teaching strategies.

Pilot team teachers who taught a CRU also had more minutes of teacher-led activities, as defined by the Special Strategies Observation System (SSOS) data collection instrument, than another group teaching a CRU, less time spent on off-task behaviors than both comparison teachers and other non-pilot team teachers using a CRU, and less time spent on student-led activities (as defined in this instrument) than in some classrooms where teachers were not using a CRU. These findings are perhaps not unexpected; the CRUs were highly planned and included many very specific components. Therefore, one could reasonably expect that there would be more time spent on teacher-led activities and less time for student-led activities, as well as less student time spent off task. Additionally, the observational findings suggest that the CRUs were implemented as they were designed. This is additional evidence to support our conclusion that teachers' participation in the full treatment improved the quality of their instruction.

Based on this study, we conclude that pilot school teachers who did not receive any component of the intervention (i.e., non-pilot teachers not teaching a CRU) have the most room to improve in terms of using an appropriate level of instruction and incentives for learning in their classroom instruction. Moreover, regarding the principles of good instruction generally and culturally responsive instruction specifically, those teachers again have the most room to improve—even more so than comparison teachers.

Finally, we conclude that the pilot schools intervention was powerful in terms of creating the positive, culturally responsive learning environment that was observed in the PTT group teachers' classrooms.

Question 6. ***To what degree does having a teacher in the full-treatment or partial-treatment group affect student engagement?***

- PTT group teachers had more success than other groups in engaging students in interactive instruction. Students in PTT classrooms were more often engaged in

interactive instruction, less often working alone, and less often social or uninvolved.

- PTT group teachers demonstrated better use of class time.

- PTT teachers who taught a CRU kept more students engaged than other groups.

In this research, we found that pilot team teachers who were teaching the unit were least likely to have students off task. Additionally, this group of teachers had the highest percentage of students engaged in interactive instruction. Students in the PTT group classes were involved in markedly higher amounts of interactive instruction than any other group (75%), which is at least 14% higher than the next highest group. Students in the PTT classrooms were also less often working alone than students in any other group. The volume of SSOS observational data, which is equivalent to at least 30 days of classroom instruction, allows us to make strong conclusions regarding this finding. Interactive instruction involvement rates above 50% have been associated with highly effective regular classroom instruction (Stallings, 1980). Interactive instruction is also indicative of active teaching, student controlled discourse, and small group instruction—three of the nine principles of culturally responsive instruction. Although all groups in the study were successful in using more than 50% of their classroom time for interactive instruction, no other groups were as successful as PTT teachers. The teaching strategies provided to the full-treatment group in workshops and team meetings, as well as the lessons in the CRU, were based on the premise that students learn more when they interact with teachers, with aides, or with peers on subjects related to the lesson. Therefore, we conclude that the full application of the intervention was successful in strengthening teachers' ability to design and teach lessons that incorporate interactive strategies and highly effective instruction.

At the same time, students in PTT group classes were involved in markedly lower amounts of off-task behavior such as social or uninvolved activities (3% of students) and "waiting" (1% of students). This would indicate that students in PTT classrooms were engaged in appropriate activities assigned by the teacher. This finding has significant implications for student learning and leads to the conclusions that students in the PTT classrooms had greater opportunities to learn and were receiving more appropriate instruction. Because the intervention was designed to maximize student learning, this finding points to its effectiveness in enhancing teachers' skills in using appropriate instruction—which can translate to more culturally responsive and differentiated instruction.

Likewise, only 5% percent of the PTT group's students were observed as out of the room, the second lowest percentage for the five groups. Generally, when students are out of the classroom they are not engaged in learning. Therefore, this finding translates into differences in opportunity to learn for students. Time out of the room, combined with the instructional time lost by some groups due to off-task behavior during class time, can have a serious negative effect on student learning. The intervention was successful,

particularly for the PTT group, in helping teachers make effective use of classroom time, ergo providing opportunities for increasing student achievement.

Notably, the PTT group teachers had the most effective use of instructional time of all the groups, including comparison teachers. The importance of time on task has long been documented as essential to student learning. On average, PTT teachers and students spent 9.45 fewer minutes per hour engaged in "off task" activities in their classrooms when compared to other groups. Projected to a full year, this amounts to a savings of 16.11 days of instructional time. This finding has strong implications for schools and districts about how effectively instructional time is used and what types of learning activities are most engaging for students. It also leads us to conclude that the intervention was effective in helping teachers improve their use of instructional time. The fact that PTT teachers taught a CRU that was purposefully designed to engage and challenge students, in addition to learning about and practicing culturally responsive teaching strategies throughout the year, led to their significantly greater ability than other groups to engage students and keep them on the learning tasks.

Question 7. ***Over the course of this project, how do educators' interpretations of the achievement gap change?***

- There is congruence between the findings of the Interview Design process and other findings in the current research.

- The Pilot Schools Project intervention made an impact on teachers' behaviors and attitudes regarding components of culturally responsive instruction for their students. For instance, responses to the Interview Design questions indicate that pilot school faculty members believe that respecting students' cultures is important for teaching African American students. Likewise, pilot teachers perceive that their schools have made significant improvements in anti-discriminatory teaching practices and responsive pedagogy throughout the course of the project. Teachers' behaviors also indicate that they have been successful in creating a culturally responsive learning environment (e.g., using culturally mediated instruction, using multiracial materials, encouraging student-controlled classroom discourse).

- Participants in the Pilot Schools intervention showed growth over time in their understanding of (1) the impact of culture and ethnicity on teaching and learning in classrooms and (2) the value of culturally relevant instruction in narrowing the achievement gap.

The findings of our research show that educators' interpretations of the achievement gap changed in substantive ways. Some teachers in the initial 2001 focus groups took the position that their responsibility was to treat all children the same, not to see color. The implication of such statements is that to see color would be to expect less from children because one has noticed their color, rather than that to see color would be to create the opportunity to use differences to enhance instruction. It was this unvoiced

implication that the project was intended to address, and the project was successful in this respect, we conclude.

The fact that the results of the Interview Design process conducted during the concluding project workshop shows growth over time in participants' understanding of the impact of culture and ethnicity on teaching and learning in classrooms and the value of culturally relevant instruction in narrowing the achievement gap gives evidence that the project achieved much of its intent. This success is further verified by classroom observations using the SOSS that show project team members exhibiting both a higher quality of instruction and greater adherence to the principles of culturally relevant instruction than their non-team colleagues. Interview Design responses also indicate project participants believe they have made progress in anti-discriminatory teaching practices, while initial focus group responses did not suggest the possibility that discrimination in classroom or schooling practices was an issue that needed addressing. Also, the fact that the AEL MSCI, an instrument completed by all project school faculties, showed growth in anti-discriminatory teaching and responsive pedagogy gives evidence that positive change in attitudes extended beyond school teams to the broader faculties. This is very encouraging news and helps to establish the spread of the findings from the two quantitative data collection methods.

Overall Conclusions

Overall, the pilot schools intervention has a positive effect on teachers' beliefs, perceptions, and behaviors about the value of culturally responsive instruction and the role it may play in improving student achievement and narrowing the achievement gap. Also, the more involvement teachers had with the intervention (e.g., being a member of the pilot team, teaching the CRU), the greater the adherence to the principles of culturally responsive instruction as presented and facilitated in this project. Thus, the intervention was more successful at progressive levels (i.e., the greater or more intense the involvement, the greater the adherence to or application of the principles of culturally responsive instruction).

RECOMMENDATIONS

The following recommendations are based on the results of the pilot schools project research.

- For the culturally responsive teaching intervention to be most effective, teachers must receive the full treatment, which in this project included the following:

  - a skilled facilitator knowledgeable about the impact of culture, ethnicity, and socioeconomic status on teaching and learning and knowledgeable about culturally responsive instruction

  - regular team meetings led by a skilled facilitator that included learning about and discussing how culture, ethnicity, and socioeconomic status impact teaching and learning; designing lessons that exemplify the principles of culturally responsive instruction; reflecting on and discussing lesson delivery and student response to lessons

  - participating in workshops on culturally responsive teaching and the experience of actually teaching culturally responsive curriculum units

- To experience the full effect of a culturally responsive teaching intervention, teachers need continuing assistance from a skilled facilitator. The facilitator could be a person from a school, district, or other agency who has received appropriate training in culturally responsive instructional strategies.

- For schools to get the maximum benefit from this intervention, implementation throughout the whole school is necessary.

- Culturally responsive curriculum units should be offered to teachers in conjunction with the appropriate context and training (e.g., professional development, workshops, ongoing technical assistance). Offering these units absent of such support does not result in the most effective teaching and does not produce the desired results in the classroom.

- Based on the findings of the possible amount of instructional time lost by students due to off-task behavior in classrooms that did not participate in the full treatment, schools should consider implementing this process as one way to decrease student time off task, especially for African American students. Likewise, schools should consider using culturally responsive instruction as one way to increase student engagement and interactive instruction.

- Schools looking to increase their capacity to improve and develop as continuously improving learning communities could adopt this model as one method for achieving those goals.

- Any implementation of this model should include the collection and examination of student achievement data and other student data both before and after implementation of the intervention.

- Culturally responsive curriculum units, geared to state content standards, should be developed as exemplars for all grade levels and multiple subject areas.

- Any future implementations of this model should be researched to determine whether results achieved in this study are replicated elsewhere.

REFERENCES

Cowley, K. S., Meehan, M. L., Finch, N., Chadwick, K., Howley, C., Riffle, J., et al. (2002). *Comprehensive evaluation of the Kentucky extended school services program.* Charleston, WV: AEL.

Cowley, K. S., Riffle, M. J. S., Howley, C. W., Voelkel, S., & Ermolov, L. (2004). *Field test of the AEL Measure of Academic Supportiveness and Climate (AEL MASC) student instrument.* Charleston, WV: AEL.

Evertson, C., & Burry, J. A. (1989). Capturing classroom context: The observation system as lens for assessment. *Journal of Personnel Evaluation in Education, 2*, 297-320.

Meehan, M. L., Cowley, K. S., Craig, J. R., Balow, N., & Childers, R. D. (2002). *AEL Continuous School Improvement Questionnaire: User manual and technical report.* Charleston, WV: AEL.

Nesselrodt, P. S, & Schaffer, E. C. (1993, April). *The ISERP programme: A revised classroom observation instrument.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Nesselrodt, P. S, & Schaffer, E. C. (2000a). *External evaluation of Kentucky's extended school services, spring 2000: Phase 1 – final report.* Carlisle, PA: Dickinson College.

Nesselrodt, P. S, & Schaffer, E. C. (2000b). *External evaluation of Kentucky's extended school services, spring 2000: Phase 1 – final report, part 2.* Carlisle, PA: Dickinson College.

Riffle, M. J. S., Howley, C. W., & Ermolov, L. D. (2004). *Measure of School Capacity for Improvement (MSCI): Early field test findings.* Charleston, WV: AEL.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Stallings, J. (1980). Allocated learning time revisited: Or beyond time on task. *Educational Researcher, 9*(11), 11-16.

Sullivan, D. K., & Meehan, M. L. (1983, April). *The development, pilot test, and field test of an instrument for the systematic observation of classroom interruptions.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

West Virginia Department of Education (WVDE). (2004). *WESTEST: West Virginia Educational Standards Test.* Retrieved November 30, 2004, from http://westest.k12.wv.us.