Measuring Treatment Integrity: Testing a Multiple-Component, Multiple-Method Intervention Implementation Evaluation Model

Merrill L. Meehan, AEL
Chandra L. Wood, AEL
Georgia K. Hughes, AEL
Kimberly S. Cowley, AEL
Jody A. Thompson,
AEL Minority Research Fellow

AEL
1031 Quarrier Street
Charleston, WV 25301

Paper presented at Evaluation, 2004, the 18th Annual Conference of the
American Evaluation Association
Atlanta, Georgia
November 3-6, 2004

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.

Information about AEL projects, programs, and services is available by writing, calling, or visiting AEL's Web site.

Post Office Box 1348
Charleston, West Virginia 25325-1348
304-347-0400
800-624-9120
304-347-0487 fax
aelinfo@ael.org
www.ael.org

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

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The issue of treatment fidelity has been a concern in the field of evaluation research (Sechrest, West, Phillips, Redner, & Yeaton, 1979). Nearly three decades ago, Cook and Campbell (1975) outlined four types of validity that may influence treatment outcomes. Defining these validity measures has since prompted researchers to closely examine potential threats within the context of program evaluation. Related to Cook and Campbell's concept of construct validity, Sechrest and associates (1979) identified two complications that could further impact evaluation research. These complications involve the strength and integrity of treatment, and their associated implications on construct validity. Shadish, Cook, & Campbell (2002) define construct validity as "making inferences from the sampling particulars of a study to the higher-order constructs they represent" (p. 65).

The *strength* of a proposed treatment is easily illustrated in the field of medicine. For instance, a physician administers varying levels of a drug in hopes to elicit a desired outcome. Assessing the strength of a treatment to varying conditions, however, is not so much a high priority among other fields of research (Sechrest, West, Phillips, Redner, & Yeaton, 1979). Consequently, this lack of concern threatens the ability to draw higher-order implications.

The extent to which the proposed treatment was delivered as planned relates to Sechrest and associates' (1979) second complication, *integrity* of treatment. This additional threat to construct validity is a result of implementation failure, typically produced by the administration of the treatment. While this failure may result from an incompetent researcher, Shadish, Cook, and Campbell (2002) add that research participants can contaminate validity by diffusing the treatment. Such erroneous implementation further complicates study results and prevents linkage between theory and outcome.

While the assessment of all types of validity has grown in evaluation research over the past decades, Sechrest and associates (1979) argue that some types of validity are emphasized more often than others, specifically, internal and external validity. Measures of internal and external validity are critical in supporting outcomes; but so, too, is the linkage between theory and a proposed treatment (Sechrest, West, Phillips, Redner, & Yeaton, 1979). They posit that higher-order implications cannot be made before first assessing the strength and integrity of treatment.

Literature Review

Juvenile offenders are subjects that have been studied regarding fidelity of their treatment. Researchers studying fidelity of treatment in the rehabilitation of juvenile offenders found that fidelity decreased when multisystemic therapy (MST) experts did not provide oversight to community mental health settings (Henggeler, Melton, Brondine, Scherer, & Hanley, 1997). When MST therapy was not emphasized, there was an increased variability in the treatment protocol. When experts emphasized MST, there was a 47% reduction of repeat offenses.

A plethora of research regarding treatment fidelity also has been conducted in the area of mental health. Many of those who have conducted such research have focused on Assertive Community Treatment (ACT). Researchers examining 50 programs that used ACT found that programs with high fidelity scores also had high scores on staffing continuity, specialist staffing, and the engagement of staff/retention of clients (Teague, Bond, & Drake, 1998). Hiring and recruiting staff was a more frequent occurrence in ACT programs that had weak fidelity scores.

McGrew, Bond, Dietzen, and Salyers (1994) also studied fidelity of program implementation of ACT programs. Treatment fidelity was defined as "conformity with prescribed elements and absence of non-prescribed elements" (p. 670). These researchers found that ACT experts believed that a team approach by therapists and case workers was beneficial in assisting clients, team members should assist clients in obtaining basic needs and entitlements, and that increased client functioning is pertinent to the client's success. Strong program fidelity was correlated with reduction in client days hospitalized in an institution.

Since trainers and facilitators are key to the success of a program's implementation, how they are recruited is significant. Trainers of a life skills training program for students were less likely to train teachers in the curriculum if they were directed by supervisors to be a trainer (Hahn, Noland, Rayens, Christie, 2002). Researchers of this study also found that only 49% of the trained teachers implemented the program one or more times. The teachers and trainers stated that they had difficulty in implementing the program because of job duties, planning lessons, and other time constraints.

Manual-based training and treatment fidelity has been studied by researchers also. In evaluating various psychotherapy studies, Miller and Binder (2002) found that manual-based training promotes adherence to specific treatments. However, trainers and therapists have an impact as to the fidelity of the treatment. Different therapists or trainers deliver the treatments at different levels of potency.

A three-step process for fidelity measurement. Mowbray, Holter, Teague, and Bybee (2003) provide further support for the assessment of treatment fidelity. First, study of treatment fidelity can determine whether outcomes are attributed to successful treatment intervention methods, or the result of implementation failure. In community-based organizations, sites within programs sometimes deviate from the intended intervention in order to accommodate their own environmental context, thus hindering treatment consistency and fidelity. In addition, assessment of fidelity can strengthen evaluation results by accounting for confounding variables that often impede intervention of a treatment.

Based on review of research incorporating measures of fidelity, Mowbray and associates (2003) have developed a three-step process for assessing fidelity criteria. The first step involves identifying and defining the criteria necessary for successful implementation of a treatment. Criteria frequently cited include measures of treatment dosage, content of the intervention, and staff involvement. Methods to develop these criteria include conducting literature reviews, researching best practices, and adapting existing framework proven effective by programs delivering similar treatment.

Following the identification of critical procedures, criteria are then measured to assess the degree to which treatment was implemented. Qualitative methods of data collection are often quantified to simplify measure of the last phase of the model. Preferred methods of measuring criteria include observations, interviews, and video recordings; self-reports of implementing treatment were perceived as less valid compared to other methods.

The last phase of Mowbray and associates' (2003) model involved assessing the reliability and validity of fidelity criteria. Methods of assessment include measuring inter-rater reliability, conducting divergent evaluations between interventions perceived to be different, and measuring convergent validity to ascertain similarities across various methods of data collection, i.e., comparing interview responses with program documents.

The researchers further indicate that current practices of fidelity assessment often lack the breadth to comprehensively measure fidelity criteria (Mowbray, Holter, Teague, & Bybee, 2003). For instance, they assert "evaluators need to examine the structure of fidelity measures and consider presenting subscores on important but significantly different critical components" (p.333). In addition, less structural approaches often need constant revision and are easily misinterpreted. Mowbray and associates (2003) recommend the use of a two-stage approach to counteract invalid assessments of fidelity criteria.

The Concerns-Based Adoption Model (CBAM). For more than 30 years, the Concerns-Based Adoption Model (CBAM) has been a major framework for the study of the installation, use, and adoption of innovations/programs. The CBAM framework has been employed mainly in the education field but recently has been used in the field of medicine. The CBAM is a useful framework for the study of individuals and groups involved in the change process. The CBAM concepts and their associated measures can be used for describing, assessing, and facilitating change in schools, colleges, and other organizations. The initial CBAM research was conducted at the R & D Center for Teacher Education at the University of Texas at Austin. CBAM researchers developed their foundation document in 1973 (Hall, Wallace, & Dossert) and embarked on an extended series of investigations to study the implementation of education innovations (Hall & Hord, 1987). Over the past three decades, the use of the CBAM to study the implementation of innovations/programs has been reported in numerous research reports, papers, evaluations, chapters, and four books (Hall & Hord, 1987; Hord, 1987; Hord, Rutherford, Huling-Austin, & Hall, 1987; Hall & Hord, 2001).

Simply stated, the CBAM is a research-based system for studying participants' changes, in terms of their concerns and behaviors, as they are involved in implementing an innovation/program. The CBAM researchers have identified, described, and developed instrumentation for the three major diagnostic dimensions of (1) Stages of Concern, (2) Levels of Use, and (3) Innovation Configurations.

For the first dimension of the model, CBAM researchers identified seven Stages of Concern (SoC) that individuals experience during the implementation of an innovation/program. These seven stages are Awareness, Information, Personal, Management, Consequence, Collaboration, and Refocusing (Hord, 1987). Researchers confirmed that there can be developmental movement through all stages, but especially so for the first four stages. They

discovered that the first four stages are highly predictable which means a participant's earlier concerns must be lowered in intensity before later concerns increase in intensity (Hall & Hord, 1987).

The second major dimension in the CBAM trio is Levels of Use (LoU). With this second dimension, CBAM researchers identified eight distinct LoU regarding an individual's implementation of the innovation/program. These eight different implementation levels focus on a person's behaviors and skills with respect to their use of the innovation/program. The names of these eight LoU are Nonuse, Orientation, Preparation, Mechanical Use, Routine, Refinement, Integration, and Renewal. Through an assessment of their behaviors and skills with the innovation/program under study, participants are said to be at one of the eight levels. This assessment is made through a personal interview by a trained and certified LoU interviewer (Hall & Hord, 2001). Attempts to assess LoU via paper and pencil surveys have failed, and Hall and Hord (2001) state clearly that it is not feasible.

The third CBAM dimension is Innovation Configurations (IC). This dimension specifically addresses what the innovation/program actually is. Whereas the SoC and LoU both focus on the participant (concerns about and use of the innovation), the IC focuses on the innovation/program itself (Hall & Hord, 2001). Innovation Configurations represent the operational patterns of the program that result from the implementation by different participants in different contexts. To discover the patterns of implementation, an Innovation Configuration Checklist (ICC) is developed. The ICC is comprised of the critical components of the innovation. Then, each critical component is divided further into the different ways the component can be implemented. These different ways to implement each component are called variations. Each participant is placed in one variation of each component in the ICC based on data provided in an interview. Practical examples of ICCs are provided by Hord, Rutherford, Huling-Austin, and Hall (1987).

Meehan and his associates have been employing the three CBAM tools, in various arrangements, in the completion of numerous evaluations in education for more than 20 years. For example, in an evaluation of the implementation of three primarily hands-on mathematics projects in elementary schools in selected counties in two states using the LoU interview tool, Meehan (1992) found that 94% of the classroom teachers were using the specific mathematics project at least at the Mechanical Use or Routine levels (p. 10). This finding was viewed as a positive outcome for the first year project. Another example of the use of CBAM tools was in the three-year evaluation of special NSF-funded professional development summer institute for earth sciences education in West Virginia (Meehan, Hambrick, & Cowley, 1995). Over the course of the comprehensive evaluation of this program, a total of 17 evaluation reports were produced, including participant observations, surveys, event evaluations, and the final evaluation. Of those 17 evaluation reports, 8 employed one or more of the CBAM tools. Hall and Hord (2001, p. 222-223) quote a passage from the final evaluation report of the earth science evaluation in which the AEL evaluators reflect over all the tools and methodologies employed in the 17 evaluation reports and conclude that, "The evaluation concluded that the combination IC Map and the LoU interviews was the most successful aspect/procedure employed in this evaluation" (p. 222). The citation continues with an explanation of five reasons for the selection of the IC Map/LoU interviews as the most successful evaluation procedure by AEL evaluators,

including the personal contact with participants, the efficiency of the dual-purpose interview, the clarity with which the major components of the program were described, the report of how the earth science program actually was being used in classrooms, and the helpful interaction of the project directors with evaluators that was required to make critical decisions about the LoU interviews and the IC Map.

A comprehensive assessment model. In efforts to comprehensively assess treatment fidelity, Lichstein, Riedel, and Grieve (1994) developed a three-component model. Delivery assessment inspects how the treatment was delivered and the factors that enhanced or hindered delivery of the treatment. Receipt assessment examines whether the intended audience was able to understand the information that was provided to them. Enactment assessment examines whether the clients' receipt of the treatment will effect them; enactment looks at whether clients adhere to the directives and change behaviors.

Lichstein, Riedel, and Grieve (1994) examined the number of research articles over a one-year period that included their three components of treatment fidelity. Forty-two studies across two clinical research journals were selected for inclusion in the study. An article's focus on treatment fidelity was evaluated simply by its promotion and/or assessment of each component. Surprisingly, only 1 out of the 42 articles addressed encouragement and assessment of all three components of treatment fidelity in their research. Twenty-one percent of the studies performed inducement of treatment delivery, receipt, and enactment, while only 10% performed assessment of all three components.

Such a void of attention on measures of treatment fidelity across existing literature further prompted Lichstein, Riedel, and Grieve (1994) to address the issues of delivery deficits and receipt deficits. Delivery deficits of treatment were said to occur "when the actual treatment delivered differs from the intended treatment by virtue of additions or deletions" (p. 11). For instance, an undedicated therapist may not fully implement a treatment that is necessary for producing an intended behavior. Consequently, treatment delivery should be assessed with this deficit in mind. Receipt deficits result "from either therapist or patient shortcomings, or from poor communication between the two for which they share responsibility" (p.12). Contrary to deficits in delivery, receipt deficits can result from either therapist or patient. An intervention that is delivered as intended does not always infer receipt of the treatment.

Lichstein and associates (1994) further dissect their model of treatment fidelity to propose various methods of assessment. First, assessment of treatment delivery must identify components of the proposed treatment, as well as components of associated treatments that are deemed unacceptable (Lichstein, Riedel, & Grieve, 1994). In terms of clinical research, recorded tapes of therapy sessions are a favored method to assess the extent to which treatment was delivered. Lichstein and associates (1994) adapted an instrument that scores criteria of treatment delivery observed on tape, as well as those procedures that corrupt the proposed treatment, i.e., leakage measures. Second, assessment of treatment receipt is determined by the extent to which "the subject was influenced by the manipulation in the manner which the experimenter wished" (p.17). Compared to delivery assessment, direct observation of treatment receipt is consistently cited as a preferred method of fidelity assessment. The last component, treatment enactment, is argued to be more difficult to assess than the former two components. Here, mixed methods of

assessment are more common, including self-reports of treatment application, and interviews of individuals who have opportunities to observe a patient's enactment of treatment. While such indirect methods of assessment are measurable, Lichstein and associates (1994) argue, "direct assessment of treatment enactment outside the therapy session is, of course, preferred and satisfies our curiosity about its level of occurrence" (p. 19).

Related to the aforementioned Lichstein and associates' model of assessing treatment fidelity, Shadish, Cook, and Campbell (2002) propose a similar model of fidelity implementation. Their model simulates the components of delivery and receipt, yet offers the distinction of renaming the third component "adherence," (p. 316) as opposed to enactment.

The MAACK Pilot Schools Project

The Kanawha County School (KCS) System in West Virginia has the largest percentage of African-American students—10%—in the state. However, only 34% of these students scored above the 50th percentile in the basic skills on the Stanford 9 in 2001 (AEL, n.d). KCS officials initially asked an Appalachia Education Laboratory, Inc. (AEL) staff member for assistance in improving the academic achievement of African-American students. AEL staff members, working with interested KCS and community members, created the Maximizing the Achievement of African-American Children in Kanawha County (MAACK) program to assist in the academic success of African-American and low income students. The two components of the MAACK program were the pilot schools project and the community initiative project. This study deals with the first component.

The KCS central office staff chose four schools (two elementary, one middle, and one high school) in which to pilot the new project. At these four sites, facilitators from AEL were to assist each school's MAACK teams by introducing culturally responsive instruction to the pilot school team members. The team members were asked to implement the principles of culturally responsive instruction in their classes and then to disseminate the information learned to other staff members in their school. The objectives of the MAACK pilot schools project were:

1. Improve academic performance of all students, particularly African-American students, many who are low socio-economic status, receiving free or reduced lunch.

2. Improve academic environment/school climate for all students, particularly for African-American students, many of whom are low SES, receiving free or reduced lunch.

3. Build the capacity and commitment of pilot schools personnel to learn and improve their professional practices.

The MAACK pilot schools project began in the 2001-2002 school year. The first two years involved AEL facilitators working with the schools to address the achievement gap problem. Much of this work was awareness, analyses of student achievement data, and co-development work. These first years set the stage for the more formal implementation of an intervention designed to address the achievement gap. The formal study of the intervention in

the MAACK schools was conducted in the 2003-2004 school year, which is the timeframe for this evaluation of the implementation of the intervention.

The MAACK project facilitators were to plan bi-monthly meetings with their pilot school team members and to record what transpired at the meetings. To assist in planning and implementing these school team meetings, AEL facilitators developed 10 guidelines to follow when presenting Culturally Responsive Instruction (CRI, see Appendix A). In essence, these CRI guidelines were the "manual" for the delivery of MAACK pilot schools project.

The facilitators adopted the principles of CRI from the Knowledge Loom, which was developed by the Education Alliance at Brown University. The nine CRI principles are as follows:

- Communication of High Expectations – The message that students can be successful should be communicated by the entire school.

- Active Teaching Methods – Curriculum should be developed to promote student communication.

- Teachers as Facilitator – The teacher should be guide and facilitate instruction in the classroom as well as teach.

- Positive Perspectives on Parents and Families of Culturally and Linguistically Diverse Students – Students, parents, and the community should have a relationship with the school and continuous dialogue should take place.

- Cultural Sensitivity – The teachers learn about the cultures in their classrooms and shape instruction accordingly.

- Reshaping the Curriculum – The background of the students shapes the curriculum.

- Culturally Mediated Instruction–Language, values, and beliefs of other cultures are intertwined in the curriculum.

- Student-Controlled Classroom Discourse – Teachers allow students to control part of the lesson.

- Small Group Instruction and Academically Related Discourse – Students are placed in small groups and work together on problems or assignments, so that they can share knowledge (The Education Alliance, n.d.).

Purpose of Evaluation

To assess the extent of the treatment fidelity of the MAACK pilot schools project, an evaluation of the AEL facilitators' work with the pilot school team members, an assessment of the teachers' receipt of the MAACK intervention, and an assessment of their adherence to the MAACK principles was undertaken.

The purpose of this study was to assess the delivery, receipt, and adherence of the CRI treatment implemented in the four Kanawha County schools. In creating a model for assessing treatment implementation, Lichstein, Riedel, and Grieve (1994) posit that there are three components in their model. Delivery assessment inspects how the treatment was delivered and the factors that enhanced or hindered delivery of the treatment. Receipt assessment examines whether the intended audience was able to understand the information that was provided to them. Adherence assessment examines whether the clients' receipt of the treatment will effect them; enactment looks at whether clients adhere to the directives and change behaviors. This study was designed to provide a viable test of the evaluation of an educational intervention employing a set of multiple components and multiple methods. These various components and methods for assessing the integrity of the intervention were suggested by Lichstein, Riedel, and Grove (1994).

INTERVENTION DELIVERY ASSESSMENT

Methods

As described above, the MAACK facilitators agreed to record the results of their bi-monthly pilot school team meetings in journal entries. These journals documented the facilitators' observations about the CRI principles introduced and the discussion between the pilot team members. These meeting note journal entries were stored in the AEL data warehouse. To learn whether the facilitators were faithful to the treatment method they collaboratively developed to implement in the MAACK pilot schools, their journal entries were examined and given a score for the CRI principles introduced. The facilitators also were interviewed individually by the evaluator to learn more about what transpired while facilitating the MAACK project.

Participants

Four AEL facilitators conducted the MAACK Pilot Schools project. One facilitator was assigned to each of the four Kanawha County schools. The project was implemented in two elementary schools, one middle school, and one high school, each having larger populations of African American and low income students than other schools in the system. These schools comprised a feeder school system (elementary to middle to high school). Two of the facilitators were full-time AEL employees, and the other two were consultants to the project (one of whom had been employed by AEL in the recent past). All of the facilitators had experience working as K-12 teachers and all were females. Each MAACK project facilitator worked with her assigned school through the intervention period (i.e., the 2003-2004 school year).

Journal Assessment Instrument

Each facilitator agreed to complete and submit journal entries from the project team meetings they conducted in their schools. The evaluator inspected the journals to determine if they followed the nine CRI principles they adopted when they conceptualized training the MAACK team members. The evaluator created a delivery assessment rating sheet, the AEL Program Delivery Assessment Form (AEL PDAF), to assess the journal entries for the principles that were introduced and discussed. For example, Student-Controlled Classroom Discourse was a principle that the facilitators planned to discuss with team members. For each CRI principle on the AEL PDAF, the facilitators' journals were rated on each of the following items: description of principle, examples of principle in the classroom, additional resources provided, supporting research given, principle reviewed, lesson plan reviewed, and student engagement discussed. This section of the rating form was called "Part 1." See Appendix B for a copy of the AEL PDAF.

For each item under the principle, the delivery of the CRI in the journal entry was read, judged, and rated by the evaluator. The ratings were: 0 if the item was *Not discussed at all*, 1 if the item was *Partially discussed*, and 2 if the item was *Fully discussed*. A facilitator's delivery of the CRI intervention was given a score of 1, for instance, if the journal indicated that they provided additional resources for Active Teaching Methods, but did not discuss the principle with their MAACK pilot school team members. An example of a facilitator receiving a score of 2 on an item would be that the facilitator wrote about the lesson plans reviewed and the discussion that took place amongst team members.

Each facilitator's project delivery journals also were given points if she assisted MAACK team members with presenting CRI to the entire school and if she assisted team members in developing a plan for continuing professional discussions with other faculty regarding CRI. These two items also were rated, but the scoring was different. If the items were not met, the score was 0, if partially met the score was 7, and fully met, the score was 14. Since the facilitators decided that they would introduce the principles to the entire school and design professional development sessions, the evaluator felt these items were just as important as the nine principles in terms of the delivery of the MAACK intervention. For each of the nine principles with each item under them, the most a facilitator could score for each CRI principle was 14. The possible point range for each facilitator's intervention delivery assessment was 0 to 154.

A "leakage scale" also was created for assessing the delivery of each of the nine CRI principles. This scale was created to provide a method of recognizing and assessing off-topic discussions during the pilot team meetings. It was labeled "Part 2" of the AEL PDAF. The three items under each of the principles were: "teachers discussed off-topic issues," "principal discussed off-topic issues," and "facilitator discussed off-topic issues." The items were rated 0 for *None*, 1 for *Partial*, and 2 for *Full Leakage*. Partial leakage, for instance, would be awarded when an off-topic issue was discussed, but the facilitator or other team member reminded the team that they need to be on task. An example of "full leakage" would be that a team member dominated the entire meeting and no one reminded the team member that he/she needed to be on task. The leakage score (if any) was subtracted from the Part 1 delivery assessment score. The range of points that could be deducted from first score was 0 to 54.

Interviews

As a supplement to the examination of the journal entries, the facilitators were interviewed individually to obtain a better understanding as to what took place during the MAACK Pilot Schools project in each school. A phenomenological interview is one way to learn how the facilitators make meaning of their experiences (Rossman & Rallis, 2002). The evaluator created an interview protocol with fourteen questions. To ensure that the questions were appropriate and comprehensive, they were reviewed by AEL's Director of Evaluation. The revised interview protocol was resubmitted to him and, based on that review, the facilitator interview protocol was revised again and copied for use.

The evaluator contacted the four MAACK facilitators by phone or by e-mail to schedule the interviews. Each interview was tape recorded and each averaged 30 minutes long. Next, the interview tapes were transcribed by administrative assistants and the transcripts were provided to the evaluator for analysis. First, responses were placed in evaluator-created categories. The evaluator looked for themes among the categories provided by the facilitators and recorded what those themes were. If the researcher was unsure about a comment by a facilitator in the transcript, clarification was obtained by listening to the original tape recording. Next, patterns of the emerging themes were identified and labeled (Posovac & Carey, 2003).

Delivery Assessment Results

Journal Assessments

The MAACK project facilitators and their school team members decided the dates and times of their meetings during the school year. The number of journals from each facilitator reflected these dates and what was discussed at these meetings. The MAACK facilitator with the most journals submitted eleven. The least number of journals submitted by a facilitator was three. It should be noted that the trainer with the most journals had two journals that reflected her work with a reading program in her school (thus, off-intervention). None of the facilitators' journals reflected that all nine Culturally Responsive Instruction (CRI) principles were discussed with team members.

Table 1 displays the results of MAACK facilitator's intervention delivery assessment based on the analysis of the facilitators' journal entry.

Table 1: MAACK Facilitators' Initial Program Delivery Assessment Scores

Facilitator	Number of Journal Entries	Delivery Score*	Leakage Score	Final Score
1	9	53	4	49
2	3	42	1	41
3	8	77	2	75
4	11	80	3	77

*The possible score range was from 0 to 154.

The score of each facilitator's delivery assessment minus the leakage scale was: facilitator 1 had a score of 49, facilitator 2 had a score of 41, facilitator 3 had a score of 75, and facilitator 4 had a score of 77. The maximum possible score was 154. Thus, the four

facilitator's scores were very close to (n = 2) to or well below (n = 2) the midpoint of the possible delivery assessment score.

Figure 1 is a graphic portrayal of the facilitators' initial program delivery assessment scores. Based on the possible point range of 0 to 154 points, the four delivery assessment score shows two to be located very close to the midpoint of 78 points, while the other two scores are in the second quarter of possible points—one just two points above the quarter mark. If the four quarters were given word descriptors representing differing levels of delivery of the interventions, from top to the lowest quarters, they would be *High*, *Moderate*, *Low*, and *Very Low*. Applying these descriptors to the facilitators' scores, all four were in the *low* delivery level, although two of those were very close to being in the *Moderate* category.

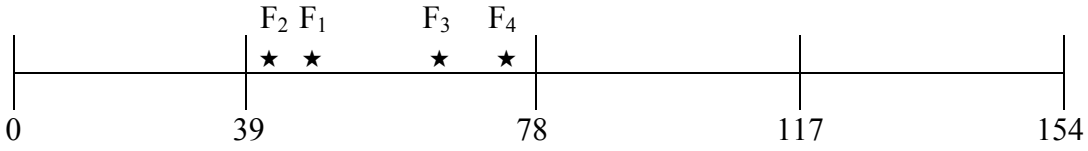

Figure 1: Graphic Portrayal of the MAACK Facilitators' Initial Program Delivery Assessment Scores

All of the facilitators introduced the Communication of High Expectations principle to their team members. Facilitator 1's journals reflected that she introduced Active Teaching Methods, Teacher as Facilitator, Positive Perspectives on Parents and Families, Student-Controlled Classroom Discourse, and Small Group Instruction to her pilot school team members. Facilitator 2's journals reflected that only Student-Controlled Classroom Discourse and Small Group Instruction were introduced in her school. Facilitator 3 introduced all of the principles except Cultural Sensitivity and Culturally Mediated Instruction. The journals of facilitator 4 indicated that she discussed all of the principles except Reshaping the Curriculum and Positive Perspectives on Families.

Interview Results

The interviews with the facilitators indicated that the facilitators did not receive formal training in CRI. The facilitators conducted planning meetings in which they all brought CRI materials to the meeting to share with other facilitators. All facilitators indicated that they either read relevant material from the website or found articles related to cultural responsiveness. One facilitator stated that a few of the nine principles could not be introduced as lesson plans such as Positive Perspectives on Parents and Family. This facilitator opined that this activity needs to be undertaken by the entire school and community.

Relationships with principals and team members. The MAACK facilitators were asked about their relationships with pilot school principals and team members. Facilitators 1 and 2 stated that they had "wonderful" relationships with the principals at their respective schools. Facilitator 2 commented:

> My relationship with the principal has been excellent all along. I really didn't know her prior to the beginning of the project, but we seemed to develop and immediate rapport and trust. In fact, the whole team was very receptive.

Facilitators 3 and 4 stated they had difficult relationships with the principals at their school. Facilitator 4 commented that the principal of her school was willing to work with the MAACK Project. However, the principal was new to the school and was still learning about her school. Facilitator 4 commented:

> I will have to say that I have never felt like I have been disappointed with the relationship with the principal and there are a number of reasons for that. The other three principals were part of the dialogue sessions that began this work between the district and the community. The principal I was working with was not a part of those conversations. She was a brand new principal and she was invited by the district superintendent to be one of the four MAACK [pilot project] schools because her school had the demographics that fit the profile and she acknowledged at one time that she did not feel she had a choice, although, it was posed as voluntary... she did not hear it as voluntary.

Facilitator 3 stated that the principal she worked with was not open to AEL staff entering her school. The principal told facilitator 3 that her school was already practicing culturally responsive principles. The principal stated that her teachers were "doing what they're supposed to." Facilitator 3 gave her opinion of the principal's attitude regarding the MAACK Project:

> She [the principal] never wanted the program in her school at all. She said, "I'm going to let you do your job, but we really don't need it." We tried everything, and that's when I asked her to let me observe some teachers and see and I kept doing that. She made me welcome after that, but none of the ideas were that important for me to reiterate, you know, they [the team members] could just go on [move forward with the meeting].

All of the facilitators commented that they had a pleasant working experience with the pilot team members at their respective schools. The facilitators kept in contact with team members and assisted those who were teachers in their classrooms. Facilitators 2 and 3 also worked with reading groups within their schools. Though facilitator 3 had a good relationship with her team members, she also said the team members were sometimes uncomfortable when the principal participated in the team meetings because the principal was not always receptive to the MAACK project taking place in her school.

Presenting the CRI principles to teams and faculty. The facilitators were asked which CRI principles were least difficult for the team members to comprehend. Three facilitators stated that Small Group Instruction and Active Learning principles were least difficult. Two facilitators also commented that the High Expectations principle was not difficult for team members to

understand. All of the facilitators observed that many of the teachers on their pilot teams already were practicing a few of the CRI principles. Facilitator 2 said:

> At the beginning, some of the teachers were in fact doing mostly large group instruction and teacher controlled classroom…you know everybody sits in their seats and listens to the teachers talk. Teachers on the team were like that. The one that I know that they got, well no, I take that back, I thought was most powerful is Student-Controlled Discourse and that is one I did see develop.

Facilitators stated the principles that were the most difficult to understand were Cultural Sensitivity and Culturally Mediated Instruction. Facilitator 4 said that Student-Controlled Discourse was difficult for teachers in her group to comprehend because the teachers wanted to be in control of all instruction. Facilitator 3 stated that Reshaping the Curriculum was difficult because teachers did not want to research additional materials. Facilitator 1 said that it was difficult for pilot school teams to understand the CRI principles because a few of the principles are abstract and not concrete. She gave an example of how the CRI principles could be difficult to understand:

> And the same thing with Reshaping the Curriculum…what does that mean if I'm a math teacher? I mean that's where you get all of these bizarre iterations of things. For example, one of the lesson plans had to do with the math teacher. She did some kind of game that was indigenous to some tribe in Africa about network; she [the teacher] felt that was reshaping the curriculum. Well, to me, that wasn't reshaping the curriculum and being culturally sensitive, that was just being bizarre. I mean, I just didn't get it.

One step that the facilitators agreed upon in introducing CRI to their respective schools was to work with team members in presenting the principles to other faculty members. The team members were to plan ways in which the materials would be introduced. However, facilitators 1 and 4 did not have time to work with the faculty because of time constraints. Facilitator 3 asked the principal to allow the team members to make a presentation to faculty. The principal allowed a teacher on the team to explain the High Expectations principle to the faculty senate.

Overall experience working with the MAACK pilot team. All of the facilitators provided feedback during their interviews on the positive aspects of introducing the CRI principles to their respective school. Facilitator 3 said she enjoyed having the opportunity to work with African American children and to help them be successful. The teachers at this facilitator's school never said no when the facilitator offered to help the teachers in the classroom. Facilitator 2 stated that she enjoyed seeing the changes in the classrooms and within the school. Facilitator 4 observed:

> And seeing changes, I mean they weren't as big and dramatic as I had wanted to see, but I think that we have seen some changes and it's not that something exists that never existed before. It's that the things that existed before in some classrooms and with some teachers has been validated and

> sort of understood to be something that needs to spread and not just idiosyncratic to those teachers.

The facilitators also provided their opinions on the improvements that might be made in facilitating the MAACK Project. Time was a major concern during the school year. The facilitators met with their teams before or after school. All of the facilitators felt that they did not have enough time to introduce the principles thoroughly. The team meetings were scheduled to last from an hour to an hour and fifteen minutes. Yet, a few times, facilitator 1 observed, her team meetings lasted two hours.

Another improvement that the facilitators said could be made was the number of principles introduced. Facilitator 4 stated that they [the facilitators] were collapsing the nine principles into five for the next iteration of the CRI intervention in schools. Facilitator 3 commented that when training teachers or teams, facilitators should start with the principles team members they might have difficulty understanding. She said:

> But some of the ones [principles] like where they have children do lessons, be involved, I think if we would start there and let them do it, we would catch them right away. It would also help our discipline. Yeah, if we would start there with the ones like Active Teaching, it's simple you know. But the other ones, like Small Group is okay, but Student-Controlled Classroom Discourse…we would start there and see what students would do. That way, you learn their culture; you learn what's happening at home, the parental involvement and all of that, right in one principle.

Recalculating Delivery Assessment Scores Based on Interviews

Using the data provided in the interviews, facilitators' delivery assessment scores were recalculated. All of the facilitators stated in their interviews that all principles were presented to their respective team members. When asked which principles she was not able to introduce to team members, facilitator 1 said, "I think we got through all of them." She also said that she did not have her team members present lesson plans on Positive Perspective on Parents and Families. Nevertheless, facilitator 1 did relay her experience in the interview discussing Active Teaching Methods, Cultural Sensitivity, and Reshaping the Curriculum with the teams members and the lesson plans that went with those principles in recalibrating her initial score she was given four points for each principle that was mentioned because she introduced the principles to her group and discussed the lesson plans for these principles. Her recalculated score for the overall delivery assessment is 61.

Facilitator 2 stated that she presented on all of the principles in her pilot team meetings. In her interview, facilitator 2 discussed the difficulty on presenting on Culturally Mediated Instruction and Reshaping the Curriculum. During her team meeting, team members talked about what lesson plans would look like and how to engage students in learning, especially if their background is different than the teacher's. Based on this information from the interview,

facilitator 2 was given six points for introducing Culturally Mediated Instruction and Reshaping the Curriculum principles, reviewing lesson plans, and discussing student engagement with team members. Thus, facilitator 2's recalculated score for overall delivery is 53.

Facilitator 3 commented in her interview that she did not introduce Culturally Mediated Instruction and Cultural Sensitivity because she did not have time to discuss those principles. When facilitator 3's delivery score was recalculated, she was given three points for discussing lesson plans with team members (Active Teaching Methods and Reshaping the Curriculum). However, during the interview, facilitator 3 stated "when they [the teachers] would write lesson plans, I would go back and talk with them about it; not in a meeting, but personally." Facilitator 3 received an additional two points each for the five principles in which she originally did not receive points for their lesson plans. Since she reviewed the lesson plans with the individual teachers personally, her recalculated score is 88.

Facilitator 4 stated that she introduced all of the principles, but there was overlap in presenting the principles. In the interview, she talked about presenting to her team members Positive Perspectives on Parents and Family. She asked facilitator 3 to come to her school to assist her with the presentation. She did not do a lesson plan on the principle; however, the facilitators and the team members did talk about the relationship between parents and the school staff. Facilitator 4 commented:

> I was thinking about the most difficult [principle]. The last one we did was Positive Perspectives on Parents and Families and I asked [facilitator 3] to come because she had lived in that neighborhood and had taught in that areas for many years...so she could speak on behalf of the community...to talk about the fact that many African-American Parents [felt] that [school name] was a school to stay away from.

Facilitator 4's recalculated score is 85. She was given eight points for introducing Positive Perspectives on Parents and Families, for giving additional resources and supporting research, and for discussing how it affects student learning and engagement.

Figure 2 is a graphic portrayal of the facilitators' recalculated MAACK program delivery assessment scores. The delivery assessment scores were recalibrated based on the intervention delivery information provided to the evaluator in the personal interviews. This program delivery information was not provided in the facilitators' journal entries. On the same 0 to 154 point range, all four scores moved up the scale. These recalibrated delivery assessment scores are more evenly distributed on the continuum. Using the same set of descriptors for the four quarters described above, the results of the recalibration now show two scores in the *Low* category and two scores in the *Moderate* category of the program delivery.

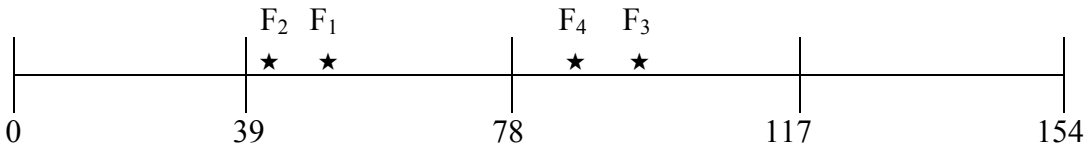

Figure 2: Graphic Portrayal of the MAACK Facilitators' Recalculated Program Delivery Assessment Scores

INTERVENTION RECEIPT ASSESSMENT

Because the MAACK Pilot Schools intervention called for facilitators to spend a great deal of time delivering training and information about culturally responsive instructional strategies and culturally responsive lesson plans, researchers assessed implementers' (i.e., school staff implementing the intervention) receipt of these two training components. To this end, AEL researchers developed the AEL Intervention Receipt Assessment (AEL IRA), a brief, two-item instrument to measure the receipt aspect of the DRA model (see Appendix C). The first item asked respondents to explain their understanding of the components of culturally responsive lesson plans; the second item asked respondents to describe their understanding of culturally responsive instructional strategies as presented in the MAACK Pilot Schools project. Both items on the AEL IRA were open-ended because AEL researchers wanted to capture the most accurate data possible about participants' understanding of the intervention they received.

AEL IRA Scoring Rubric

AEL staff developed a two-part rubric to score participants' responses to the AEL IRA (see Appendix D). The two parts of the rubric assess the extent to which respondents explained the components associated with each of the two items on the AEL IRA. In collaboration with MAACK Pilot School Project developers, the researchers determined that there were 12 important components of culturally responsive lesson plans; these 12 components were used as the criteria for the first part of the evaluation rubric. The nine components of culturally responsive instruction detailed in the MAACK Pilot Schools intervention framework served as the criteria for the second part of the scoring rubric.

AEL IRA responses were rated using a three-point scale indicating the extent to which the response explains each criterion. The ratings for each criterion on the AEL IRA scoring rubric range from *0*, indicating that the criterion was not mentioned, to *2*, indicating that the criterion was well or fully explained. Responses in which a criterion was mentioned but not explained were rated *1* for that criterion. Points were assigned according to the rating (i.e., 0 points for a *0* rating) and were tallied for each part of the rubric. A total of 24 points was possible for the first part of the AEL IRA assessing the 12 lesson plan components; a total of 18 points was possible for the second part assessing the nine components of culturally responsive instruction. Thus, total scores on the AEL IRA could range from 0 to 42.

Receipt Data Collection

AEL staff administered the AEL IRA during a MAACK Pilot Schools training workshop held on May 6, 2004. Because the workshop itself might have influenced participants' memories about the components of the intervention, the AEL IRA was administered at the beginning of the workshop in order to obtain the most accurate reflection of participants' understanding and knowledge of the intervention. All participants at the workshop, excluding the school system superintendent, were asked to respond to the instrument. An AEL researcher explained the directions for the instrument and monitored participants as they responded. Participants needed

approximately 20 minutes to complete the task. When all participants had finished responding to the AEL IRA, an AEL researcher collected the completed instruments and submitted them for scoring.

Receipt Data Analysis

Three AEL researchers independently rated participants' responses to the AEL IRA using the scoring rubric described previously. AEL IRA responses were scored in a holistic manner. Although the instrument and the scoring rubric are divided into two parts, raters accepted responses from both parts of the instrument when rating each criterion. For instance, a respondent might have mentioned "Teacher as Facilitator" in the first part of the AEL IRA even though it is not a criterion for Part 1. Raters, however, accepted that response and scored it as a response to Part 2, where "Teacher as Facilitator" is a criterion (see Appendix D).

After completing independent ratings of each AEL IRA, researchers met to discuss the responses, their ratings, and discrepancies in ratings. Researchers attempted to resolve as many discrepancies in ratings as possible. Agreement among the three raters ranged from 66.7% to 100% after the raters met to resolve discrepancies. For most of the AEL IRA instruments, agreement among the three raters ranged from 81.0% to 100%; in only five instances did interrater agreement fall below 80%. The three raters' total scores for each AEL IRA were then averaged, and the mean of the three scores was used as the final total score.

Receipt Assessment Results

Twenty seven AEL IRA forms were completed and collected at the spring 2004 workshop. Included among those 27 instruments were five from KCS central office staff and two that included no information identifying the respondents' school affiliations. Because the primary focus of this study is to examine receipt of the intervention by those who were to implement the culturally responsive instruction in their classrooms, the AEL research team decided to exclude the central office staff and the two respondents who could not be identified from the final analyses. Thus, the final sample of respondents included 20 staff members from the four pilot schools

Implementers' scores on the AEL IRA ranged from 2.00 to 20.33 in a possible range of 0 to 42.00. Most scores (n = 13) ranged between 2.00 and 6.00; only three respondents received scores greater than 10.00 on the AEL IRA. The mean score for all implementers was 6.38 (SD = 4.32). Table 2 presents basic descriptive information for the implementers' AEL IRA scores by the facilitators' code numbers.

Table 2: Implementers' AEL IRA Scores by Facilitator's Code Number

		Range			
Facilitator	***n***	**Low**	**High**	**Mean**	**SD**
1	3	3.00	5.00	4.11	1.02
2	12	3.00	20.33	7.97	4.86
3	2	2.00	3.67	2.83	1.18
4	3	3.00	7.33	4.67	2.33
All	*20*	*2.00*	*20.33*	*6.38*	*4.32*

The total range of scores on the AEL IRA (0 to 42) can be divided into fourths to classify the receipt of the intervention from *Very Low* to *High*. The guidelines listed in Table 3 can be used to classify the extent of receipt.

Table 3: Guidelines Classifying Extent of Receipt of the Intervention

Classification	Range of AEL IRA Scores
Very Low	0.0 – 10.5
Low	10.6 – 21.0
Moderate	21.1 – 31.5
High	31.6 – 42.0

The majority participants at the spring 2004 workshop received AEL IRA scores that indicate Very Low receipt of the intervention (n = 18). Only two respondents received scores greater than 10.6, indicating Low receipt of the intervention. No respondents could be classified as Moderate or High in receipt of the MAACK Pilot Schools project intervention. Table 4 presents these data by facilitator code number.

Table 4: Participants' Receipt of the Intervention by Facilitator Code Number

Facilitator	Very Low	Low	Moderate	High
1	3	0	0	0
2	10	2	0	0
3	2	0	0	0
4	3	0	0	0
All	18	2	0	0

INTERVENTION ADHERENCE ASSESSMENT

Adherence Component Data Collection

Two different classroom observation instruments were employed to collect data on the teachers regarding their adherence to the principles of culturally responsive instruction, which was the target intervention. Those two instruments were the QAIT Assessment of Classroom and the Classroom Environment and Resources Checklist (CERC). Each will be explained briefly below.

The QAIT Assessment of Classroom (QAIT) is best described as a moderate- and high-inference, simple coding, rating device. The QAIT was developed by Slavin and is based on his research (1987 and 1989) and research of Carroll (1963 and 1989). QAIT stands for Quality of Instruction, Appropriate Level of Instruction, Incentives for Students, and Use of Time. This two-page observation instrument contains 40 items grouped under four major categories. Each item contains 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 classroom observation session, which was 56 minutes or one class period, whichever was longer.

The Classroom Environment and Resources Checklist (CERC) 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 multi-racial materials, posted assignments, etc. 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, like the QAIT.

The QAIT and CERC were two instruments of a three-part classroom observation system. Observers trained in the classroom observation system collected the data during the 2003-2004 school year. Each observation was expected to last 56 minutes, although not all did. A total of 315 classroom observations, or 14,333 minutes, were completed across all the treatment groups. By design, teachers were observed before, during their implementation of the culturally responsive unit, and after the teaching of the unit. A total of 54 different classrooms in the eight study schools were observed approximately six times each (before, during, and after).

Adherence Component Index Development

In order to determine the adherence to the principles of culturally responsive instruction (CRI) of those observed pilot team teachers, some of whom were teaching one of the units and some of whom were not teaching the unit, an index score was generated from those data points that most closely aligned to any of the nine specific components of culturally relevant instruction. See Table 5 for a depiction of the nine components, along with the QAIT and CERC observation instrument items most closely aligned with each component. 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 group of items per component was then generated as a new variable, which was then analyzed by grouping.

Table 5: Culturally Responsive Instruction 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.

Adherence Assessment Results

Adherence index. This score reflects the adherence to the principles of culturally responsive instruction for those pilot team teachers observed who were or were not teaching one of the CRI units. Twenty-two items from the QAIT and CERC classroom observation instruments that most closely aligned to one of the nine specific components of culturally responsive instruction 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 5 to see which items comprise each component score).

There was a total of six pilot team teachers observed who were teaching one of the culturally responsive units, and a total of seven pilot team teachers observed who were not teaching one of the units. All 13 of these teachers had received the training (intervention) on culturally responsive instruction. The number of observations for these 13 teachers ranged from four (one of the PTT* teachers) to seven (one of the PTNT teachers). Most teachers (69%) were observed six times (twice before a unit was taught, twice while a unit was being taught, and twice after a unit was taught).

Theoretically, the Z scores have no floor or ceiling values because the extremes at either end of the distribution go into infinity. As a practical reference for Z scores, four standard deviations above and below the mean of 50 would yield scores ranging from 10 to 90; ninety nine point nine percent of the scores would fall between these scores. The adherence scores for the two pilot team groups of teachers' CRI were from -23.2 to 129.8. Thus, four pilot team teachers (only one teaching the CRI unit) had adherence scores under the mean Z value and the remaining nine pilot team teachers (including four not teaching the CRI unit) had adherence scores above the mean Z value. For the PTT teachers, their average adherence score ranged from 49.6 to 111.8, with a mean of 81.23 and a standard deviation of 23.70. The adherence scores for the PTNT teacher group revealed much more variance, ranging from -23.2 to 129.8, with a mean of 55.31 and a standard deviation of 51.44.

A one-way ANOVA was generated to determine whether statistically significant differences existed among the five groups of teachers for this adherence variable; the ANOVA was significant ($F(4, 310) = 8.21$, $p < .05$). Since the Levene test of homogeneity of variance was significant, an unequal 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. Figure 3 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.

* PTT - pilot teacher teaching a culturally responsive unit; PTNT - pilot teacher not teaching a culturally responsive unit; NPTT - non-pilot teacher teaching a culturally responsive unit; NPTNT - non-pilot teacher not teaching a culturally responsive unit; Comp. - comparison school.

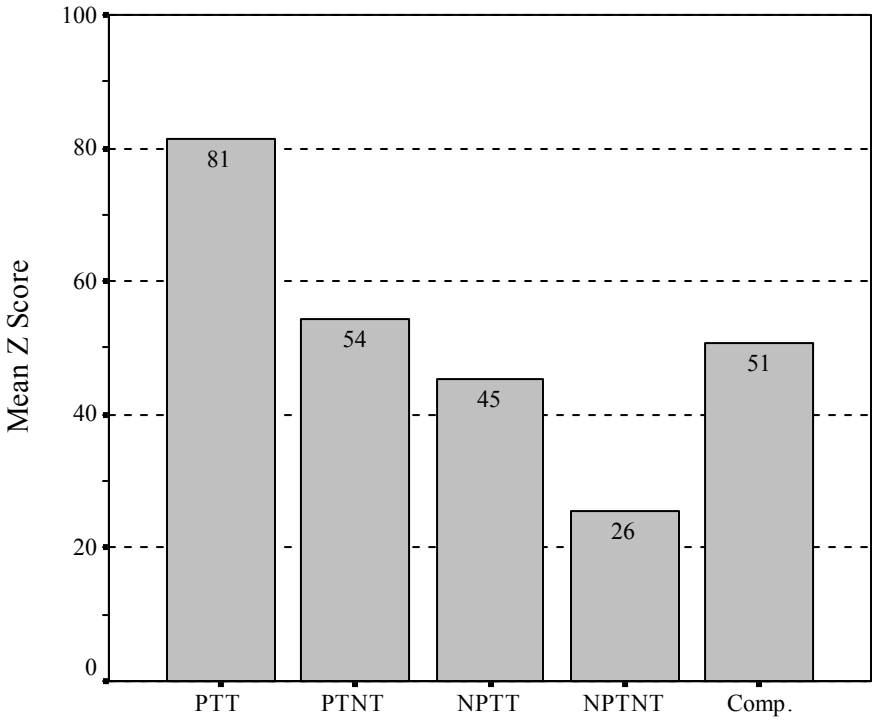

Figure 3: SSOS Culturally Responsive Adherence Index Score by Groupings*

* PTT - pilot teacher teaching a culturally responsive unit; PTNT - pilot teacher not teaching a culturally responsive unit; NPTT - non-pilot teacher teaching a culturally responsive unit; NPTNT - non-pilot teacher not teaching a culturally responsive unit; Comp. - comparison school.

CONCLUSIONS AND RECOMMENDATIONS

The primary purpose of the current study was to test the multi-component DRA model in an evaluation of an education intervention. Several conclusions about the intervention itself can be drawn from the research presented in this paper. However, the current paper focuses on conclusions and lessons learned about the DRA model and its implementation. Strengths of the model and suggested improvements are outlined in this section.

Strengths

Multiple raters and multiple observers were employed in two of the three components of the model (Receipt and Adherence). The raters and observers were able to communicate with each other and resolve questions or discrepancies throughout the evaluation process. Further, the raters were able to call attention to responses or behaviors that might have been overlooked or wrongly evaluated by a single rater. Evaluation team members, thus, were very comfortable with the final ratings and findings for the components of the model for which multiple raters were involved. Only one person, however, assigned ratings and leakage scores for the journals and interviews; because other members of the evaluation team were not able to corroborate ratings, they were slightly less comfortable with the final ratings and findings for the Delivery component. Evaluation team members considered the use of multiple raters a strength of the current study. Future studies employing the DRA model should make use of multiple raters for all components of the model.

Because the integrity of intervention implementation can be affected by the introduction of unintended elements, evaluators for the current education program thought it very important to include an assessment of those unintended, off-intervention elements. Thus, a leakage score was included to account for the introduction of features that were not designed to be part of the intervention delivery. The evaluator for the Delivery component found that there were several instances of off-intervention discussion during the MAACK pilot team meetings. Including the leakage score thus allowed evaluators to account for the "pollution" of the intervention delivery. Future applications of the DRA model should continue to include leakage scores or some similar method of accounting for intervention interference.

When evaluators began to design the AEL IRA scoring rubric, they consulted the intervention facilitators to get their feedback about the comprehensiveness and appropriateness of the criteria. Gaining collaboration and cooperation from the facilitators was an important step in gaining ownership in the evaluation from important stakeholders and is consistent with the Joint Committee's standards for program evaluation (Joint Committee, 1994). Evaluations using the DRA model should be sure to secure collaboration from program developers, implementers, and other stakeholders to ensure that assessments are complete and accurately reflect the program as it was designed for implementation.

Another strength of the current study was that evaluators had a clear starting point from which to design the evaluation and its assessment instruments. Evaluators were given access to the materials facilitators used to implement the intervention (e.g., guidelines for MAACK pilot

team meetings, principles of culturally responsive instruction). Often, aspects of interventions are nebulous or not clearly delineated. However, the MAACK pilot schools intervention for the 2003-2004 school year had specific guidelines and benchmarks for its implementation. This specificity made designing assessment instruments (e.g., AEL PDAF) easier because the criteria for implementation already were defined. Although evaluators often do not have control over how well program implementation guidelines are defined, evaluators should work with program developers and implementers to define the "it" and its benchmarks before the program is implemented.

In addition to assessing multiple components, the model under study used multiple methods to assess the integrity of intervention implementation. Evaluators used facilitators' implementation journals and in-depth interviews to assess Delivery, a knowledge assessment to gauge Receipt, and behavioral observations to assess Adherence. A particular strength of the model in this study was the use of the behavioral observations by highly skilled observers. Such observations probably gave a more accurate reflection of adherence than a self-report assessment would have done. Designers of the current study also were particularly pleased with the use of multiple methods within the evaluation of the Delivery component of the DRA model. Through the interviews with the project facilitators, evaluators were able to address information gaps in the journals and, in some cases, the absence of those journals. Use of multiple methods throughout the model and within the individual components of the model enabled evaluators to gain a complete assessment of the intervention's implementation. Future use of the DRA model should continue to make use of multiple methods of assessment. Additionally, future studies employing this model should use multiple methods for each individual component.

The five member evaluation team included four on-site AEL researchers and one off-site research fellow. The off-site research fellow conducted the interviews with the project facilitators, all of whom have continuing relationships with the on-site AEL researchers. Having the off-site researcher conduct the interviews with the facilitators allowed the on-site evaluators to maintain neutral relationships with them. Although communicating with the off-site researcher was sometimes slow or difficult, the benefit of having an off-site researcher conduct potentially sensitive interviews was a strong point of the study. Future studies in which evaluators might have continuing relationships with program implementers should consider employing off-site consultants to evaluate the Delivery component. Doing so can help avoid skepticism, possible hostility, and hurt feelings.

Although the DRA model was tested as a summative evaluation, the model holds much promise as a tool for formative evaluations. The ability to provide the facilitators with formative information about the delivery and receipt of the intervention might have helped them in adjusting MAACK pilot team meetings to better meet the needs of pilot team members. Providing similar data about receipt and adherence to pilot team teachers might have enabled them to adjust their attention to and application of the principles of culturally responsive instruction. Adjustments could be made to adapt the DRA model as an effective tool for formative evaluations. Future studies are needed to test the utility of the DRA model in formative evaluations.

Areas for Improvement

The DRA model as tested in this study has many strengths and shows much promise as another tool for both formative and summative evaluations. There are, however, several areas in the model, as tested, that could be improved. As stated previously, multiple raters are needed for each component to ensure accurate ratings. Other suggested improvements are outlined below.

Although multiple assessment methods were employed for the Delivery component and multiple types of observations (e.g., behavioral, environmental) were employed in the Adherence component, evaluators used only the AEL IRA to assess receipt of the intervention. Although evaluators originally planned to rate teacher work samples (i.e., lesson plans), they were unable to do so because of an inadequate sample size. Evaluators were provided access to lesson plans teachers produced as part of the MAACK pilot schools project; however, not all teachers at all schools completed lesson plans for each of the nine CRI principles. For instance, only three teachers completed lesson plans for Reshaping the Curriculum, and only four teachers submitted lesson plans for Positive Perspectives on Parents and Families. Thus, the sample size for the lesson plans was insufficient to evaluate teachers' work samples. Future studies should be sure to employ multiple assessment methods for each of the model's three components to ensure comprehensive evaluation.

Including the leakage score to compensate for the introduction of off-intervention elements was a strength of the study. However, evaluators did not adjust for leakage based on the results of the facilitator interviews. Thus, although leakage scores were applied to the journal scores, they were not applied to ratings of the interviews. This oversight is a weakness of the current study, and future applications of the model should account for leakage in all assessments of treatment delivery.

Several factors could have impacted receipt of the intervention, such as prior knowledge of culturally responsive instructional principles or strategies. However, evaluators had no method of assessing factors that could have influenced receipt scores. Assessing participants' prior knowledge or other factors that might have influenced receipt of the intervention could be considered analogous to adjusting delivery scores for leakage. Future studies should consider addressing such influences or leakage not only in delivery but also in receipt of the intervention.

Another weakness of the current study was outside the control of the evaluators. There was a great deal of variability in the amount of data for each facilitator in the Delivery component of the model, in part because there were so few facilitators. One facilitator submitted only three journals, and another submitted 11; thus, there was a great amount of variability in the amount of data analyzed for the Delivery component. The small number of facilitators also limited the extent to which evaluators could examine the role of the facilitator in receipt of and adherence to the intervention. In other studies with a larger number of facilitators or program deliverers, evaluators might be able to examine the role of a facilitator's delivery in the other components of intervention implementation.

The model, as tested in this study, includes the three components (i.e., Delivery, Receipt, Adherence), assessed and rated separately. Although the delivery and receipt scores can be linked based on the school the facilitator worked with, those scores cannot be related to the adherence data. The primary reason for this was because some teachers were very reluctant to provide an individual identification number on the Receipt assessment instrument. There should be a way to link the scores from the three components to one another to create a composite treatment integrity score. Future studies should attempt to collect data in such a manner that delivery, receipt, and adherence scores can be assigned to each individual implementing the program.

Although facilitators were consulted about the construction of instruments and scoring rubrics, they were not informed from beginning about the use of this model as part of the evaluation. This, in part, was due to the fact that the opportunity to test the novel model design presented itself after the intervention had been underway for several months. Evaluators, thus, were unable to inform the facilitators about the nature of this form of evaluation and ensure their collaboration from the outset of the intervention. If the design had been ready to test at the time the intervention began, evaluators would have been able to secure the full cooperation of the facilitators at that time and thus avoid some of the skepticism facilitators felt about the journal assessments and interviews. Evaluators also might have been able to measure the concerns of the change facilitators, part of the CBAM model, which could have enriched the evaluation. Future studies employing this model should consider assessing the concerns of change facilitators and incorporating other aspects of the CBAM model, such as the Levels of Use interview technique to assess the Receipt component.

Another improvement to the model would be an assessment of "dosage strength." That is, intervention recipients often receive different levels, or doses, of an intervention; evaluators employing the DRA model should attempt to gauge the strength of recipients' doses of the intervention. In the current study, pilot team teachers were divided into two groups: one group taught a culturally responsive curriculum unit (CRU), and one group that did not teach a CRU. The group of pilot team teachers who taught the CRU might have received a higher dose of the intervention because they were more involved in the intervention than the other group of pilot team teachers. Measuring dosage strength of intervention delivery could enhance the assessment of both the receipt and adherence components of the model.

Although the model and instruments as specifically applied in this evaluation might not be of use for some interventions, the general DRA model can be adapted and applied in myriad situations. This model shows promise for many types of interventions and could be of great utility in the field of evaluation; as discussed earlier, the DRA model could be beneficial when used in a formative or a summative capacity. Such additions to evaluation methodology are needed, especially in the field of education where the current trends mandate research bases for instructional programs and strategies. Given current political climate and funding requirements, comprehensive, start-to-finish evaluations, especially of educational innovations, might be well worth the efforts to test and refine the methodology. The current study is the first step and should be followed with additional studies and tests of the DRA system.

REFERENCES

AEL. (n.d.). Maximizing the Achievement of African American Children in Kanawha (MAACK) Pilot School Project. Charleston, WV: Author.

Cook, T. D., & Campbell, D. T. (1975). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunette and J. P. Campbell, *Handbook of Industrial and Organizational Research*, 223-325. Chicago, IL: Rand McNally.

Hahn, E. J., Noland, M. P., Rayens, M. K. & Christie, D. M. (2002). Efficacy of training and fidelity of implementation of the Life Skills Training Program. *Journal of School Health*, *72*, 282-287.

Hall, G. E., & Hord, S. M. (1987). *Changes in schools: Facilitating the process*. Albany, NY: State University of New York Press.

Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Needham Heights, MA: Allyn and Bacon.

Hall, G. E., Wallace, R. D., & Dossert, W. A. (1973). *A developmental conceptualization of the adoption process within educational institutions*. Austin, TX: University of Texas at Austin, R & D Center for Teacher Education.

Henggeler, S. W., Melton, G. B., Brondino, M.J., Scherer, D. G. & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, *65*, 821-833.

Hord, S. M. (1987). *Evaluating educational innovation*. London: Croom Helm.

Hord, S. M., Rutherford, W. L., Huling-Austin, L. & Hall, G. E. (1987). *Taking charge of change.* Austin, TX: Southwest Educational Development Laboratory.

Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards, 2nd Edition*. Thousand Oaks, CA: Sage Publications, Inc.

Lichstein, K. L., Riedel, B. W., Grieve, R. (1994). Fair tests of clinical trials: A treatment implementation model. *Advanced Behavior Research Theory*, *6*, 1-29.

McGrew, J. H., Bond, G. R., Dietzen, L. & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology*, *62*, 670-678.

Meehan, M. L. (1992). *1991-92 Evaluation of project H.O.M.E.: Levels of Use (LoU) interview results*. Charleston, WV: Appalachia Educational Laboratory.

Meehan, M. L., Hambrick, K. & Cowley, K. S. (1995). *Final evaluation report: Earth science in West Virginia for the twenty-first century (Rock Camp)*. Charleston, WV: Appalachia Educational Laboratory, Inc.

Miller, S. & Binder, J. L. (2002). The effects of manual-based training on treatment fidelity and outcome: A review of the literature on adult individual psychotherapy. Psychotherapy: Theory, Research, Practice, Training, *39*,184-198. Retrieved July 20, 2004 from http://gateway2.ovid.com.ezproxy.lib.vt.edu:8080/ovidweb.cgi.

Mowbray, C. T., Mark, C. H., Teague, G. B. & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24,* 315-340.

Posovac, E. J. & Carey, R. G. (2003). *Program evaluation: Methods and case studies*. Prentice Hall: Upper Saddle River, NJ.

Rossman, G. B. & Rallis, S. F. (2003). *Learning in the field: An introduction to qualitative research,(2nd edition)*. Thousand Oaks, CA: Sage Publications.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R. & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. *Evaluation Studies Review Annual, 4,* 15-35.

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

Teague, G. B., Bond, G. R., & Drake, R. E. (1998). Program fidelity in assertive community treatment: Development and use of a measure. *American Journal of Orthopsychiatry*, *68*, 216-232.

The Education Alliance. (n.d.). *The Knowledge Loom: Culturally responsive teaching*. Retrieved July 7, 2004 from http://knowledgeloom.org/practices3.jsp?location=1&bpinterid=1110&spotlighted=1110.

APPENDICES

APPENDIX A

Guidelines for Facilitating Bi-Monthly MAACK Team Meetings

Guidelines for Facilitating Bi-Monthly MAACK Team Meetings

1. Each month select one of the nine principles of culturally responsive instruction (www.knowledgeloom.org) as a focus of team discussion.

2. Provide team members with copies of the description of the principle, examples of how the principle "looks" in the classroom, additional resources they may use to implement the principle, and the supporting research (www.knowledgeloom.org).

3. Ask team members to read the material. Lead a discussion of the key points. Ask for examples of how team members may be addressing the principle.

4. Ask team members to use this principle in their classrooms during the next week or two. Ask team members to describe exactly what they will do differently. Explain that each team member will be expected to share with the group at the next meeting what they did differently and how that practice affected student learning, particularly for African American students.

5. At the next team meeting, invite team members to share their classroom experience with the new culturally responsive practice and to discuss among themselves the benefits of this practice for improving engagement and learning. (See protocol for instructional discussion and sharing student work.) Ask the teachers to share lesson plans with others, including AEL staff. They may use the journal template provided by AEL or whatever lesson plan format they are currently using in the school.

6. Discuss ways to share this information within the school and develop a plan for continuing professional discussion with other faculty. This plan may include peer observation and coaching.

7. Discuss ways in which AEL can assist the team with "rolling out" culturally responsive teaching to the whole school.

8. After the culturally responsive unit has been taught, discuss its effect on student engagement and learning, particularly for African American students.

9. After each meeting, the AEL facilitator should document in writing what occurred at the meeting. Include the number of teachers and administrators present, examples of teachers' comments on what worked with their students, questions/concerns team members may have, and other information that may affect the school's progress toward improving achievement of African American students.

10. Share your personal reflections with AEL staff regularly, use them to write quarterly report copy, and complete appropriate forms required by AEL.

APPENDIX B

AEL Program Delivery Assessment Form

Date: _____ **Facilitator's School:** _____
Rater ID: _____ **Facilitator ID:** _____

AEL Program Delivery Assessment Form

INSTRUCTIONS: Read each journal entry and assess each item for each component by circling the appropriate response number (0, 1, 2 or 0, 7, 14). Then, compute the component subtotals, the part totals, and the grand total.

Part 1: Culturally Responsive Instruction Methods

CRI Component	Items	Delivery Assessment None	Partial	Full
Communications of High Expectations	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal
Active Teaching Methods	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal
Teacher as Facilitator	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal
Positive Perspectives on Parents and Families	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal

Part 1: Culturally Responsive Instruction Methods (Continued)				
		Delivery Assessment		
CRI Component	Items	None	Partial	Full
Cultural Sensitivity	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal
Reshaping the Curriculum	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal
Culturally Mediated Instruction	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal
Student-Controlled Classroom	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed		1	2
	Student Learning and Engagement Discussed			
				Subtotal
Small Group Instruction and Academically Related Discourse	Description of Principle	0	1	2
	Examples of Principle in Classroom	0	1	2
	Additional Resources Given	0	1	2
	Supporting Research Given	0	1	2
	Principle Reviewed	0	1	2
	Lesson Plans Reviewed	0	1	2
	Student Learning and Engagement Discussed	0	1	2
				Subtotal

Part 1: Culturally Responsive Instruction Methods (Continued)				
		Delivery Assessment		
CRI Component	Items	None	Partial	Full
	Facilitator assisted MAACK team members with presenting CRI to the entire school	0	7	14
	Facilitator assisted MAACK team members in developing a plan for continuing professional discussions with other faculty	0	7	14
				Subtotal
				Part 1 Total

Part 2: Leakage Scale				
		Delivery Assessment		
CRI Component	Items	None	Partial	Full
Communication of High Expectations	Teachers discussed off-topic issues	0	1	2
	Principal discussed off-topic issues	0	1	2
	Facilitator discussed off-topic issues	0	1	2
				Subtotal
Active Teaching Methods	Teachers discussed off-topic issues	0	1	2
	Principal discussed off-topic issues	0	1	2
	Facilitator discussed off-topic issues	0	1	2
				Subtotal
Teacher as Facilitator	Teachers discussed off-topic issues	0	1	2
	Principal discussed off-topic issues	0	1	2
	Facilitator discussed off-topic issues	0	1	2
				Subtotal
Positive Perspectives on Parents and Families	Teachers discussed off-topic issues	0	1	2
	Principal discussed off-topic issues	0	1	2
	Facilitator discussed off-topic issues	0	1	2
				Subtotal
Cultural Sensitivity	Teachers discussed off-topic issues	0	1	2
	Principal discussed off-topic issues	0	1	2
	Facilitator discussed off-topic issues	0	1	2
				Subtotal

Part 2: Leakage Scale (Continued)

CRI Component	Items	Delivery Assessment None	Partial	Full	
Reshaping the Curriculum	Teachers discussed off-topic issues	0	1	2	
	Principal discussed off-topic issues	0	1	2	
	Facilitator discussed off-topic issues	0	1	2	
					Subtotal
Culturally Mediated Instruction	Teachers discussed off-topic issues	0	1	2	
	Principal discussed off-topic issues	0	1	2	
	Facilitator discussed off-topic issues	0	1	2	
					Subtotal
Student-Controlled Classroom	Teachers discussed off-topic issues	0	1	2	
	Principal discussed off-topic issues	0	1	2	
	Facilitator discussed off-topic issues	0	1	2	
					Subtotal
Small Group Instruction and Academically Related Discourse	Teachers discussed off-topic issues	0	1	2	
	Principal discussed off-topic issues	0	1	2	
	Facilitator discussed off-topic issues	0	1	2	
					Leakage Subtotal
					Part 2 Total
					GRAND TOTAL

APPENDIX C

AEL Intervention Receipt Assessment (AEL IRA) Instrument

AEL INTERVENTION RECEIPT ASSESSMENT (IRA)

School: _____ **Last four digits of your SS Number:** __ __ __ __

This instrument is designed to assess MAACK Pilot School Project team members' understanding of culturally responsive instruction as delivered by AEL facilitators. **Please respond to the following items to the best of your ability in the spaces provided.** Thank you for your time.

1. Please describe your understanding of the **components of culturally responsive instruction lesson plans** as presented in the MAACK Pilot Schools Project.

Please continue on the next page →

2. Please describe your understanding of the **components of culturally responsive teaching/instruction** as presented in the MAACK Pilot Schools Project.

APPENDIX D

AEL IRA Scoring Rubric

AEL IRA Scoring Rubric

Responses to IRA items should include references to the components listed for each item. Multiple references to the same component will only count as one reference. Use the following scale to rate responses:

0 = Not mentioned **1 = Component is mentioned, not explained** **2 = Component is well or fully explained**

	0	1	2	Comments
Part #1: Components of Culturally Responsive Lesson Plans (12 Items)				
Objective/purpose				
Standards-based				
Recruit interest in the task/ "hook"				
Communicate expectations				
Explain/clarify the task				
Modify content to connect with/include students' cultures				
Student involvement/engagement				
Provide extension/enrichment/modified activities for special student needs & interests				
Address multiple learning styles/multiple intelligences				
Provide guided practice & support for struggling students				
Assessment strategies/methods				
Reflection on student learning (after lesson is taught)				
Totals:				
Part #2: Components of Culturally Responsive Teaching/Instruction (9 Items)				
High expectations				
Active teaching methods/student engagement				
Teacher as facilitator				
Positive perspectives on parents and families of diverse students				
Cultural sensitivity				
Reshaping/modeling the curriculum to be culturally responsive/relevant				
Culturally mediated instruction				
Student controlled classroom activities				
Small group instruction/activities				
Totals:				