

Validation of the AVID Certification Self Study (CSS): A Measure of AVID Secondary Program Implementation Fidelity

ABSTRACT

Advancement Via Individual Determination (AVID) is a college preparatory system implemented in grades four through twelve with the mission of closing the achievement gap by preparing all students for college and career readiness. As AVID grew from one classroom to more than 4,000 schools in 47 states and 16 countries, it became apparent that a process for ensuring program quality and implementation fidelity was needed. The Eleven Essentials and Certification Self-Study (CSS) were developed to assist AVID schools in the implementation of AVID and provide the AVID Center with information necessary to monitor the quality, consistency, and implementation fidelity of AVID programs around the world. The purpose of this study is to evaluate the psychometric properties and validity of the CSS – an instrument designed to measure the implementation fidelity of the AVID secondary (grades six through twelve) college readiness program. Data from 2,655 sites were obtained using the CSS during the 2008-09 academic year. Results indicated that each of the scales measuring the Eleven Essentials met sufficient levels of internal consistency with the exception of the middle school rigor scale. The overall reliability findings allowed further investigation regarding the validity of the CSS. It was hypothesized that AVID sites with higher levels of implementation fidelity would demonstrate significantly higher student achievement consistent with the theoretical underpinnings of the CSS. Results indicated that schools implementing AVID at the highest levels of fidelity evidenced significantly higher student achievement across all academic and course enrollment outcomes. Recommendations for future research and possible changes to the current CSS are provided.

INTRODUCTION

The purpose of this study is to evaluate the psychometric properties and validity of the Certification Self Study (CSS) – an instrument designed to measure the implementation fidelity of the Advancement Via Individual Determination (AVID) secondary college readiness program.

Implementation fidelity is defined as "...the degree to which teachers and other program providers implement programs as intended by the program developers" (Dusenbury, Brannigan, Falco, & Hansen, 2003). Implementation fidelity has been shown to moderate the relationship between interventions and their intended outcomes (Elliott & Mihalic, 2004; Dane & Schneider, 1998). This relationship speaks to the importance of ensuring acceptable levels of fidelity prior to evaluating program effectiveness. Moreover, researchers have demonstrated that implementation fidelity affects the potency of the intended effects such that higher levels of fidelity produce more positive outcomes (Baker & Lorezetti, 2007; Noel, 2006).

AVID History

AVID was developed and first implemented by Mary Catherine Swanson in 1980 at Clairemont High School in San Diego, California. As an English teacher she quickly recognized several academic challenges that would later manifest for many of the minority, low-income, non-native English speaking students bused to this otherwise all-white, middle class campus. Her greatest fear was that academic rigor, faculty expectations, and student experiences would soon decline in reaction to the growing heterogeneity of the student body and the academic skills they brought with them. What was borne of these concerns is what we today refer to as the AVID college readiness system – a system of academic rigor and supports implemented in elementary, middle, and high schools and colleges and universities across the country. Mary Catherine's unwavering commitment and belief in the notion that exposure to rigorous curriculum coupled with appropriate academic supports produces high achieving students became the hallmark of the AVID philosophy.

History of the AVID Essentials

In the beginning, AVID was a single high school academic elective course offered during the school day to students in the academic middle who were underachieving, yet demonstrated academic potential, and had the determination to be academically successful. As AVID grew, so too did the understanding that the AVID elective alone was not going to bring about the type of change needed to support students at levels necessary for college acceptance and success – the goal and mission of AVID. Needed also, was the full support of site administrators and counselors; a commitment of resources to hire and retain college tutors and provide professional development to staff; and lastly, the collection and use of data to monitor and document program effectiveness and the success of students.

Mary Catherine began collecting data from the very outset. She kept information on her AVID students regarding the classes they completed, the test scores and grades they achieved, and the college/university to which they applied and were accepted. When she moved to the San

Diego County Office of Education to disseminate AVID to schools beyond San Diego City Schools in 1986, it was clear to her that a formal process to ensure the highest levels of implementation fidelity was needed. It was then that the “AVID Guiding Principles” were developed as a precursor to the 1996 adoption of the 10 AVID Essentials. Although maintaining a strong site team was part of the AVID philosophy from the beginning, the site team work wasn’t fully codified as “Essential 11” until 1999 completing the conceptual and structural underpinnings of the AVID secondary program (see Appendix A).

In 2001, the Center for Research, Evaluation, and Training in Education (CREATE) conducted a study to assess the efficacy of the Eleven AVID Essentials. Eight high school AVID sites in California were identified as highly effective based on AVID student achievement (academic performance, college acceptance and attendance rates) and recommendations by the California Regional Directors. Interviews with key program and school staff, tutors, and students provided data during two rounds of site visits to each site. In addition, classroom observations provided evidence regarding the implementation of each of the Eleven Essentials. The most significant finding was the high level of implementation fidelity each site evidenced relative to the AVID model. The authors wrote, “A key finding... was that each of the eight programs followed the AVID design almost to the letter.” The authors validated the efficacy and “essential” nature of each of the Eleven Essentials noting repeatedly how each component is critical to the success and sustainability of an AVID program (Guthrie & Guthrie, 2002).

History of Certification

In 1986, the certification process became more formalized as AVID sites began submitting a one page written summary of their program status, documenting their implementation of what would later become AVID’s Eleven Essentials. Certification continued to be an annual self-study by site teams who collected data on the fidelity of their AVID elective class to the AVID model and submitted a one-page summary each spring. In the Fall of 1991, the first Certification Validation visits were conducted by Mary Catherine and her team verifying the comprehensive implementation of the AVID Guiding Principles. These initial visits went beyond examining the AVID elective classroom proper and required that AVID strategies be evidenced in at least 3 of the 4 core content areas, namely; math, English, science, and social science. Sixteen high schools were validated initially, each demonstrating significant increases in AP course enrollments, completion of the CSU/UC “a-f” college entrance requirements, and college acceptance rates as a direct reflection of the implementation fidelity of their AVID programs.

As AVID grew nationally, a need arose to create a certification instrument which would increase the objectivity of the certification process and be accessible online. Thus, AVID Center piloted a new, more detailed concrete certification instrument in Texas from 2001 through 2004 under the direction of Dr. Charles Powell, Texas AVID State Director. The first goal was to create an instrument that would be less subjective, more objective, and reportable in an online format. The second goal was to create language and data points which could be used to generate more coaching conversations to guide sites toward more effective implementation of the AVID secondary program. During the pilot years in Texas, this new instrument identified concrete evidence for AVID district directors to use when implementing the certification process and

when conducting certification visits. It also generated useful data which could be disaggregated to provide insightful ways to assess the effectiveness of the AVID elective class and secondary program to meet AVID's goal of preparing more students for four-year college/university work.

Because of the increased usefulness and effectiveness of the multi-page Texas pilot certification instrument, AVID Center, via the leadership of Dr. Michele Marcus in collaboration with a task force of district and regional directors, created a national online certification instrument which could account for state differences and yet retain an objective review of AVID's Eleven Essentials. During the 2004 academic school year, AVID Center piloted this instrument nationally with selected districts and schools in California, North Carolina and Texas. Based on the results of the evaluations in 2005, the "AVID Certification Report and Self-Study Continuum" was revised and introduced as the official AVID Center secondary certification instrument beginning with the 2005 academic school year. It has been revised annually and used by AVID sites to document their implementation of the Eleven Essentials. Today, all AVID secondary sites have the annual opportunity to review their program status and submit their data online via the "AVID Certification Report and Self-Study Continuum."

The purpose of this study is to examine the reliability and validity of the Certification Self-Study which is the tool AVID secondary sites use to measure their implementation fidelity to the AVID model. This will be explored by determining the extent to which each *Indicator* is correlated or related to the overall Essential that it represents. The goal is to ensure that each *Indicator* within an Essential is measuring the same thing, albeit a little differently. This is referred to as internal reliability which is a necessary but not sufficient condition toward establishing the validity of a scale or measurement tool. It is expected that each *Indicator* representing a particular Essential will be most highly correlated with that Essential than with any other Essential. In other words, any one *Indicator* of Essential 1 will be more highly correlated with the overall rating of Essential 1 than with any other Essential.

In addition, to further establish the validity (the extent to which the CSS measures what it is intended to measure) of the CSS, analyses will be conducted testing the hypothesis that higher levels of implementation fidelity are associated with higher student achievement. If, in fact, the CSS is a valid measure of AVID implementation fidelity than sites with higher levels of program fidelity should produce more positive outcomes consistent with the intent of the AVID secondary program.

Methodology

Description of the Certification Self-Study (CSS)

The Certification Self-Study (measure of implementation fidelity) is written around the Eleven Essentials and is intended to provide valuable information guiding the program site team toward a deeper, more comprehensive implementation of AVID. Each of the Eleven Essentials is operationally defined using from three to seven *Indicators* scored across a four-point continuum of implementation fidelity: 0 = Not AVID, 1 = Meets Certification Standards, 2 =

Routine Use, and 3 = Institutionalization (see Appendix B). The Essentials continuum is adapted from Dr. Gene Hall's "Levels of Use of the Innovation" (Hall, 1975) whereby:

- **Level 0 "Not AVID"** refers to the absence of the essential components or implementation is at a level not meeting minimum certification requirements.
- **Level 1 "Meets Certification Standards"** refers to the one to three year time period required to implement the essential components of an innovative program
- **Level 2 "Routine Use"** refers to the two to four year time period required to implement an innovative program long enough whereby its essentials are replicated enough to become routine
- **Level 3 "Institutionalized"** refers to the three to five year time period required before a school can thoroughly implement the essentials with accurate fidelity to the model. "Institutionalized" is defined as "the integration of program-related change into the school or school system as a standard characteristic feature of school or district operations."

The site team scores each *Indicator* using the four-point continuum based on evidence sources they identify. After rating each *Indicator*, an overall Essential rating is determined.

Each Essential is rated along the same four-point continuum described above using the ratings of individual *Indicators*. While the number of *Indicators* may vary from one Essential to another no one Essential is conceptualized or understood to be of higher importance within the context of implementation fidelity. Rather, some Essentials are more complex and broad in their application and consequently require more *Indicators* to fully describe accurate and successful implementation. In light of this, when determining the overall rating of each Essential the same criteria is used regardless of the number of *Indicators* being evaluated.

Each Essential is rated using the following criteria:

- ***Institutionalization (Level 3)***: No more than one *Indicator* is below Level 3 (i.e., 2 out of 3, or 3 out of 4, etc. *Indicators* must be checked at Level 3); no *Indicator* may be at level 0.
- ***Routine Use (Level 2)***: No more than one *Indicator* is below Level 2; no *Indicator* may be at level 0.
- ***Meets Certification Standards (Level 1)***: No more than one *Indicator* is below Level 1; there must be a plan in place and implemented to bring that to Level 1 or higher for the following school year.

- **Not AVID (Level 0):** More than one *Indicator* is below Level 1.

The overall measure of implementation fidelity (or certification level) for a site is based on the ratings for each of the Eleven Essentials. For purposes of this study, three distinct groups of AVID secondary sites were created based on their overall certification level: “Non-Certified/Affiliate”, “Certified”, and “Demonstration”.

Non-Certified/Affiliate sites are those with lower levels of implementation fidelity with one or more of the Eleven Essentials rated as *Not AVID (Level 0)*.

Certified sites are those with minimum to moderately high levels of implementation fidelity such that all Eleven Essentials are rated *Meets Certification Standards (Level 1)* or higher.

Demonstration sites have the highest levels of implementation fidelity. Middle schools at this level must be certified for at least two consecutive years. High schools must be certified for at least three consecutive years and have at least one graduating class of seniors of twenty or more. All Demonstration sites must have all Eleven Essentials rated at *Routine Use (Level 2)* with no *Indicator* at Level 0. In addition, an AVID Regional or District Director must recommend the site as a model of high implementation fidelity which is later verified via an AVID Center onsite validation visit.

Description of Data Collection

Data were collected from 3,321 secondary sites during the spring semester of the 2008 academic year. Elementary and postsecondary sites were excluded from the analyses as the CSS does not measure implementation fidelity in these contexts. Of the 3,321 sites completing the CSS, 2,655 met the following criteria for inclusion in the analysis of the psychometric properties of the instrument: 1) Site had implemented the AVID program for at least two years and 2) The District or Regional Director and AVID Center approved the CSS. The CSS approval process requires a site’s District or Regional Director to go to the site, review the CSS and supporting evidence, and verify the validity of the data. Upon their approval and recommendation, the District or Regional Director forwards the CSS to the AVID Center for final approval and assignment of the appropriate certification status.

Analyses pertaining to the hypothesis that programs with higher levels of implementation fidelity produce higher student achievement included data from sites meeting the following criteria: 1) Site had implemented the AVID program for at least two years, 2) Both the District or Regional Director and AVID Center approved the CSS and 3) the AVID Center approved General and Senior data, where applicable. General data is required to be submitted on an annual basis by all AVID secondary sites. The AVID Site Coordinator collects and submits data regarding student demographics, socio-economic status, grade level, course enrollment, and various student outcomes including but not limited to: Algebra completion (at the middle level), Advanced Placement or International Baccalaureate course enrollment (at the high school

level), SAT or ACT participation, and college application, acceptance, and planning to attend rates. Programs report General data in the aggregate whereas Senior data is collected at the student level and is required to be submitted by all twelfth grade AVID students enrolled in the AVID elective.

Results

One purpose of the current study is to evaluate the internal consistency of each of the scales representing the Eleven Essentials. Items (or *Indicators*) within a scale determined to have sufficient internal consistency are said to be measuring the same construct, or in this context, the same Essential. Simply put, are the *Indicators* of each Essential measuring the same thing?

Cronbach's Alpha was calculated as a measure of internal consistency. The alpha is determined by the combined correlation that each scale item has to the overall scale. Cronbach's Alpha can range from 0.0 to 1.0 with .70 or higher generally regarded as sufficient in terms of demonstrating internal consistency (Cronbach, 1951).

Internal Consistency Using Cronbach's Alpha

Internal consistency of responses obtained on the Certification Self-Study or CSS was evaluated for each of the scales representing the Eleven Essentials. Participating secondary site teams rated themselves on their level of implementation of individual *Indicators* within each of the Eleven Essentials using a scale from 0 to 3 (see Appendix B). Each of the *Indicators* within an Essential is designed to measure a slightly different aspect of that Essential. For example, Essential 2 speaks to the importance that both students and staff **choose** to participate in AVID. The *Indicators* measuring this Essential are described as evidence of:

- 1) student contracts indicating parent and student desire to enroll in AVID,
- 2) AVID teachers volunteering to participate,
- 3) Site Team members volunteering to participate,
- 4) a selection process for identifying and choosing AVID elective teachers, and
- 5) a selection process for identifying and choosing Site Team members.

Theoretically, the implementation ratings of each *Indicator* within the same Essential should be similar if the site is implementing AVID as it is prescribed and the *Indicators* are measuring the same Essential. In other words, it is assumed that sites implementing AVID with fidelity to the model would afford equal levels of diligence (and consequently similar implementation ratings) to each *Indicator* particularly where there is clear inter-relatedness between them. Cronbach's Alpha is the metric that is used to measure the consistency in the ratings a site team gives to *Indicators* within any given Essential. In instances where Cronbach's Alpha is equal to or greater than .70, the group of *Indicators* measuring the particular Essential is considered to have produced a sufficient level of internal consistency.

For these analyses Essentials 4 and 9 were split in four smaller scales represented as 4a, 4b, 9a, and 9b, as each contains *Indicators* specific to High Schools, Middle Schools, or both (see Appendix C for the composition of these subscales). Results indicated that each of the scales

measuring the Eleven Essentials produced Alpha's equal to or exceeding .70 except Essential 4b - middle school rigor (see table 1).

Table 1.
Cronbach's Alpha for Each of the Scales Measuring the Eleven Essentials

<u>Essential</u>	<u>Number of Indicators</u>	<u>Number of Schools</u>	<u>Cronbach's Alpha</u>
1	3	2,655	0.82
2	5	2,655	0.87
3	3	2,655	0.70
4a [*]	4	1,397	0.71
4b ^{**}	5	1,317	0.59
5	5	2,655	0.88
6	4	2,655	0.84
7	3	2,655	0.87
8	5	2,655	0.86
9a ⁺	6	937	0.80
9b ^{**}	4	2,655	0.86
10	5	2,655	0.77
11	7	2,655	0.90

* Essential 4a contains some items specific to high school rigor that are engaged by high schools only

** Essential 4b contains some items specific to middle school rigor and are engaged by middle schools only

+ Essential 9a contains items measuring the submission and use of data by the AVID site team and two items specific to high schools with AVID seniors

** Essential 9b contains items measuring the submission and use of data by the AVID site team only

In addition to internal consistency, each of the *Indicators* was investigated in terms of its relationship to the Essential it was designed to measure and its relationship to each of the other Essentials. Item-total correlation coefficients were examined to determine if each *Indicator* representing a particular Essential was most highly correlated with that Essential. Evidence in this regard will further support the internal reliability of each Essential which is a necessary but not sufficient condition toward establishing the validity of the CSS. It is expected that each *Indicator* representing a particular Essential will be most highly correlated with that Essential than with any other. For example, it is expected that any one *Indicator* of Essential 1 will be most highly correlated with the overall rating of Essential 1 than with any other Essential.

Item-Total Correlation Analysis

The item-total correlation analysis was conducted to check whether any *Indicator* was not consistent with the rest of the *Indicators* representing the Essential being measured. This analysis was done using the Pearson Product Moment Correlation Coefficient which measures the ratings of an individual *Indicator* against the sum of the ratings of the remaining *Indicators* from that scale. In instances where the correlation between an *Indicator* and the Essential it was designed to measure is less than that of another Essential, the *Indicator* may be said to be

more reflective of the Essential to which it is most highly correlated. Furthermore, *Indicators* with item-total coefficients of .30 and below should be scrutinized as to their appropriateness and usefulness as items of this nature tend to produce more “noise” than understanding of the Essential for which they were designed to measure.

Of the fifty-three *Indicators* representing the Eleven Essentials, five were found to be more highly correlated with an Essential other than that for which it was designed to measure (see Appendix C). Results regarding each of these five *Indicators* are discussed below:

- *Indicator 3.1* is designed to measure Essential 3 which speaks to the complete implementation of the AVID elective including its offering during the school day. It speaks specifically to ensuring that the AVID Elective class is scheduled as a year-long course and is offered during the regular school day. This *Indicator* did correlate slightly higher with the overall rating of Essential 1 (.54 vs. .51), however, the practical significance of this difference does not warrant removal or repurposing it as an *Indicator* of Essential 1 – Student Selection.
- *Indicator 4b.1* is designed to measure Essential 4 which speaks to ensuring that AVID students are enrolled in a rigorous course of study that will enable them to meet requirements for university enrollment. More specifically, *Indicator 4b.1* speaks to middle school students and their enrollment in rigorous courses. This *Indicator* did correlate slightly higher with the overall rating of Essential 5 (.49 vs. .48), however, the statistical and practical significance of this difference does not warrant removal or repurposing of this *Indicator*.
- *Indicator 4b.3* is also designed to measure Essential 4 which, again, speaks to ensuring that AVID students are enrolled in rigorous courses. More specifically, *Indicator 4b.3* speaks to middle school students’ enrollment in an Algebra course, that upon successful completion, would allow them to enroll in Geometry or a higher math course. This *Indicator* did correlate slightly higher with the overall rating of Essential 9 (.23 vs. .22), however, given the very low correlation it does not warrant repurposing the *Indicator* to reflect Essential 9.
- Another measure of Essential 4, *Indicator 4b.5*, speaks to middle school students’ participation in pre-collegiate testing, namely, the PSAT, PLAN or EXPLORE. This *Indicator* correlated higher with every other Essential except 5, 6, and 7 (see Appendix C). Because this *Indicator* did not achieve a correlation rating higher than .35 with any of the Essentials, repurposing this *Indicator* to measure any other Essential is not warranted at this time.
- *Indicator 9a.5* is designed to measure Essential 9 which speaks to monitoring of student progress through the analysis and use of data. More specifically, It speaks to the proportion of students passing their state’s respective high school exit exam. This *Indicator* did correlate slightly higher with the overall rating of Essential 4 (.42 vs. .39),

however, the practical significance of this difference does not warrant repurposing it as an *Indicator* of Essential 4 – Rigorous Course Enrollment. That being said, it is not too surprising that this *Indicator* was found to be related to the Rigor Essential as students enrolling in these types of courses are more likely to be successful on their high school exit exams.

Based on the overall findings from the internal consistency and item-total correlation analyses, the scales representing each of the Eleven Essentials are determined to be sufficiently reliable to pursue investigating the validity of the CSS.

Description of Validity Analyses

A series of Analysis of Variance (ANOVA) procedures was conducted to determine if implementation fidelity is significantly related to various program characteristics and student outcomes.

A total of 2,701 middle and high schools were placed in one of three implementation groups based on CSS ratings and recommendations from their respective District or Regional Director. Of these, 215 were classified as Non-AVID middle schools, 1,057 were Certified middle schools, 62 were Demonstration middle schools, 174 were Non-AVID high schools, 1,137 were Certified high schools, and 56 were Demonstration high schools. These six groups were then collapsed into three distinct groups of implementation fidelity; Non-Certified, Certified, and Demonstration.

Consistent with Hall's (1975) conceptualization of program implementation, average length of AVID implementation was significantly related to implementation fidelity such that increased years of implementation led to higher fidelity, $F_{(2,2641)} = 38.97, p < .001$. More specifically, nearly one third (31%) of the of Non-certified sites had AVID for less than three years while more than 90% of Demonstration sites had AVID for five or more years. The average number of years of implementation for Non-certified, Certified, and Demonstration sites is 5.5, 6.2, and 9.5 years, respectively.

Student Characteristics

Prior to evaluating program effects on student outcomes, analyses were conducted to determine if student characteristics significantly differed across the three implementation groups. Theoretically, AVID programs serve the same population of students regardless of level of implementation. Essential 1 speaks to recruitment of students meeting the AVID profile which is characterized by students in the academic middle, underrepresented minorities, low income, and first generation college goers. It is expected therefore, that higher percentages of students meeting these criteria are enrolled in AVID when compared to the student body overall.

Analyses revealed that indeed, higher percentages of minority students are enrolled in AVID across all three groups. For example, among Non-certified sites 77.3% of AVID students are minority versus 71.6% at these sites overall. Among Certified and Demonstration sites the

pattern is the same, 73.5% versus 63.8% and 78.8% versus 72.8%, respectively. Significant differences in minority enrollment between the three levels of implementation fidelity were also identified, $F_{(2,2607)} = 6.33$, $p < .01$. While, Non-certified and Demonstration sites enrolled significantly higher percentages of minority students than Certified sites, nearly 3 out of 4 students in each group is of minority status.

Groups were also compared on the percentage of AVID enrolled, low socioeconomic students as defined by eligibility for free or reduced lunch. Not surprisingly, the exact same pattern emerged, namely, higher percentages of AVID students were free/reduced lunch eligible than the overall student body across all three groups; Non-Certified 52.9% vs. 48.9%, Certified 44.5% vs. 41.5%, and Demonstration 52.6% vs. 52.2%. Significant differences in free/reduced lunch eligibility between the three groups were also identified, $F_{(2,2604)} = 13.13$, $p < .001$. Non-certified and Demonstration sites enrolled significantly higher percentages of free/reduced lunch eligible students than did Certified sites with the largest difference being 8.4%.

Student Outcomes - Middle School Analyses

Non-Certified, Certified, and Demonstration sites were compared on student outcomes reflective of the AVID middle school program which speak to student access and enrollment in rigorous courses and preparation for acceptance to post secondary institutions of higher education. Thus, analyses were conducted using the following outcome variables: recommendation for high school enrollment in a college sequence of courses, proportion of students enrolled in Algebra or Honors courses, and completion of the PSAT or Explore test.

Middle school analyses were conducted using the General Data Collection information submitted by 1,334 Middle schools – 215 Non-Certified, 1,057 Certified, and 62 Demonstration.

Recommendation for a College Sequence of Courses in High School

Significant differences were found between fidelity groups in the average percentage of 8th grade students recommended for a college prep sequence of courses in high school, $F_{(2,1247)} = 8.65$, $p < .001$. Demonstration sites recommended a significantly higher proportion of students for a college prep sequence (96.2%) than Certified (88.5%) and Non-Certified (82.0%) sites. In addition, Certified sites recommended a significantly higher proportion of students for college prep in high school than Non-Certified sites, $p < .01$.

8th Grade Enrollment in an Algebra Course

Significant differences were found between fidelity groups in the average percentage of 8th grade students enrolled in Algebra, $F_{(2,1275)} = 4.21$, $p < .05$. Student enrollment in Algebra was significantly higher at Demonstration (39.3%) sites than at Certified (31.5%) and Non-Certified (34.0%) sites.

8th Grade Completion of a High School Math Course With a Grade of C or Better

Significant differences were found between fidelity groups in the average percentage of 8th grade students earning a grade of C or better in a high school level math course, $F_{(2,1192)} = 5.13$,

$p < .01$. Students enrolled in AVID at Demonstration sites were significantly more successful (65.4%) than students enrolled at Certified (52.9%) and Non-Certified (51.2%) sites.

8th Grade Enrollment in an Honors Course

Significant differences were found between fidelity groups in the average percentage of middle school students enrolled in an Honors course, $F_{(2,1027)} = 4.30$, $p < .05$. Student enrollment in an Honors course was significantly higher at Demonstration (64.7%) sites than at Certified (52.8%) and Non-Certified (47.0%) sites.

8th Grade Completion of an Honors Course With a Grade of C or Better

Significant differences were found between fidelity groups in the average percentage of 8th grade students earning a grade of C or better in an Honors course, $F_{(2,987)} = 3.95$, $p < .05$. Students enrolled in AVID at Demonstration sites were significantly more successful (54.9%) than students enrolled at Certified (48.3%) and Non-Certified (40.3%) sites. More specifically, Certified and Demonstration sites significantly differed from Non-Certified sites but did not differ from one another.

8th Grade PSAT or Explore Test Participation

Significant differences were found between fidelity groups in the average percentage of 8th grade students taking the PSAT or the Explore test, $F_{(2,1247)} = 56.48$, $p < .001$. Significantly more students at Demonstration sites took a pre-collegiate exam (54.9%) than students enrolled at Certified (48.3%) and Non-Certified (40.3%) sites. Certified site participation was also found to be significantly higher than at Non-Certified sites, $p < .01$.

Student Outcomes - High School Analyses

Groups were compared on student outcomes reflective of the AVID high school program which speak to student access and enrollment in rigorous courses and acceptance to post secondary institutions of higher education. Thus, analyses were conducted using the following outcome variables: completion of college entrance requirements, Advanced Placement or International Baccalaureate course and exam participation, SAT or ACT participation, completion of the Free Application for Federal Student Aid (FAFSA), and college application, acceptance and planning to attend rates.

High school analyses were conducted using the Senior Data Collection information submitted by 18,449 twelfth grade AVID students enrolled at 870 high schools – 174 Non-Certified, 1,137 Certified, and 56 Demonstration.

Completion of College Entrance Requirements

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* completing college entrance requirements, $F_{(2,867)} = 9.29$, $p < .001$. A significantly higher proportion of *seniors* at Demonstration sites completed college entrance requirements (93.7%) than *seniors* enrolled at Certified (88.0%) and Non-Certified (82.7%) sites. In addition, significant differences were identified between Certified and Non-Certified sites, $p < .01$.

Advanced Placement or International Baccalaureate Course Enrollment

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* taking at least one AP/IB course, $F_{(2,871)} = 13.98$, $p < .001$. A significantly higher proportion of *seniors* at Demonstration sites enrolled in at least one AP/IB course (84.6%) than *seniors* enrolled at Certified (67.3%) and Non-Certified (61.0%) sites. In addition, significant differences were identified between Certified and Non-Certified sites, $p < .05$.

Advanced Placement or International Baccalaureate Exam Participation

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* taking at least one AP/IB exam, $F_{(2,871)} = 20.21$, $p < .001$. A significantly higher proportion of *seniors* at Demonstration sites took at least one AP/IB exam (72.6%) than *seniors* enrolled at Certified (50.8%) and Non-Certified (43.9%) sites. In addition, significant differences were identified between Certified and Non-Certified sites, $p < .05$.

SAT or ACT Exam participation

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* who took either the SAT or ACT exam, $F_{(2,869)} = 13.57$, $p < .001$. A significantly higher proportion of *seniors* at Demonstration sites took either the SAT or ACT exam (96.2%) than *seniors* enrolled at Certified (84.5%) and Non-Certified (78.2%) sites. In addition, significant differences were identified between Certified and Non-Certified sites, $p < .05$.

Completion of the Free Application for Federal Student Aid (FAFSA)

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* completing the FAFSA, $F_{(2,869)} = 11.79$, $p < .001$. A significantly higher proportion of *seniors* at Demonstration sites completed the FAFSA (74.7%) than *seniors* enrolled at Certified (58.9%) and Non-Certified (56.0%) sites.

Submission of 4-Year College or University Application

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* applying to a 4-year college or university, $F_{(2,869)} = 11.37$, $p < .001$. A significantly higher proportion of *seniors* at Demonstration sites applied to a 4-year college or university (95.3%) than *seniors* enrolled at Certified (83.4%) and Non-Certified (79.0%) sites.

Acceptance to a 4-Year College or University Among Those Who Applied

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* who applied to a 4-year college or university and were accepted, $F_{(2,857)} = 3.62$, $p < .05$. A significantly higher proportion of *seniors* at Demonstration sites were accepted to a 4-year college or university (89.6%) than *seniors* enrolled at Certified (83.3%) and Non-Certified (81.5%) sites.

Percentage Planning to Attend a 4-Year College or University Among Those Who Were Accepted

Significant differences were found between fidelity groups in the average percentage of AVID *seniors* who plan to attend the 4-year college or university to which they were accepted, $F_{(2,486)}$

= 3.05, $p < .05$. A significantly higher proportion of *seniors* at Demonstration (77.9%) and Certified (74.0%) sites planned to attend the 4-year college or university to which they were accepted than Non-Certified (64.4%) sites.

Discussion

The purpose of this study was to evaluate the psychometric properties and validity of the Certification Self-Study (CSS) – an instrument designed to measure the implementation fidelity of the AVID secondary (grades six through twelve) college readiness program. Researchers have documented the importance of establishing implementation fidelity of interventions prior to investigating their validity (Elliott & Mihalic, 2004; Dane & Schneider, 1998) particularly in light of research suggesting that implementation fidelity can moderate program outcomes (Baker & Loretti, 2007).

Results indicated that each of the scales measuring the Eleven Essentials met sufficient levels of internal consistency with the exception of the middle school rigor scale. This is due in part to the low item-total correlations of two of the *Indicators* – *4b.3* and *4b.5*. *Indicator 4b.3* speaks to ensuring that at least 50% of AVID students successfully complete Algebra prior to the completion of the 8th grade. Findings suggest that this *Indicator* may not be indicative of the overall construct of middle school rigor as defined by the CSS as the correlation of this *Indicator* to the overall Essential score is very low. Similar explanation can be afforded *Indicator 4b.5* which speaks to the proportion of 8th grade students having taken either the PSAT, PLAN, or EXPLORE test. It appears that these two *Indicators* may be measuring expected student outcomes as a result of implementing AVID with fidelity as opposed to measuring the fidelity with which AVID is implemented. One might argue this a classic instance of putting the cart before the horse...

Alternative explanations speak to issues of state and local policy that influence student enrollment in Algebra in the 8th grade and/or participation in pre-collegiate testing. For instance, some districts and states prohibit the enrollment of 8th graders in an Algebra course unless certain prerequisite courses have been taken or test scores obtained. These institutional challenges, otherwise not directly related to the implementation of AVID, may have impacted CSS ratings. Similarly, fees associated with taking the PSAT, PLAN or other pre-collegiate exams may have impacted CSS ratings in ways not relevant to the fidelity with which AVID was implemented. In short, questions remain regarding the appropriateness and relevance of both of these *Indicators* as measures of AVID implementation fidelity.

It should be noted that while Essential 9 (the use of data to monitor student progress) demonstrated adequate internal reliability, *Indicators 9a.5 and 9a.6* showed similar attributes to the aforementioned *Indicators* of Essential 4. More specifically, they both evidenced the lowest item-total correlations, .39 and .41 respectively, suggesting their content may be more reflective of student outcomes associated with exposure to sound AVID programs as opposed to attributes of implementation fidelity. Upon closer inspection, both of these *Indicators* indeed reflect student outcomes, i.e. success on state mandated exit exams and completion of the senior data collection requirements, rather than program implementation attributes.

In terms of validity, it was hypothesized that AVID sites with higher levels of implementation fidelity would demonstrate significantly higher student achievement consistent with the theoretical underpinnings of the CSS. This hypothesis was supported such that schools implementing AVID at the highest levels of fidelity evidenced significantly higher student achievement across all academic and course enrollment outcomes. In other words, sites that implement AVID as it is prescribed and trained in Summer Institute and other AVID trainings can expect student outcomes to be significantly higher than under any other conditions.

In sum, the current findings are consistent with those previously cited by other researchers regarding the moderating effects of program implementation fidelity on participant outcomes. These results have significant implications with regard to the importance of program implementation monitoring, site level coaching, tracking of student outcomes, and establishing the efficacy of program outcomes in support of AVID expansion and continued funding.

It is understood that future efforts should focus on validating the CSS ratings of schools in the Certified and Non-Certified/Affiliate groups to ensure the accuracy of overall certification status. As previously indicated, the Demonstration schools' CSS ratings were the only group validated through an onsite visit. Observational verification of site-team ratings on the CSS at Certified and Non-Certified/Affiliate schools will help provide further evidence of the validity of the CSS. Lastly, given the sample size and variability evidenced among sites in the Certified group, further investigation of possible differences within this group of schools is certainly warranted.

References

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, Vol. **16**(3), 297-334.
- Dane A., Schneider B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clinical Psychology Reviews*, Vol. **18**, 23-45.
- Dusenbury L., Brannigan R., Falco M., Hansen W. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, Vol. **18**, No. 2, 237-256.
- Elliott D., Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth*, Vol. **4**, 83-86.
- Guthrie L., Guthrie GP. (2002). The magnificent eight: AVID Best Practices Study.
- Hall G., Loucks S., Rutherford W., Newlove B. (1975). Levels of use of the innovation: a framework for analyzing innovation adoption. *Journal of Teacher Education*, Vol. **26**, 52-56.
- Noel P. (2006). The impact of therapeutic case management on participation in adolescent substance abuse treatment. *American Journal of Drug and Alcohol Abuse*, Vol. **32**, 311-327.
- Thomas RE, Baker PRA, Lorenzetti D. (2007). Family-based programmes for preventing smoking by children and adolescents. *Cochrane Database of Systematic Reviews*, Issue 1. Art. No.: CD004493. DOI: 10.1002/14651858.CD004493.pub2.