

*Adherence: A More Nuanced Usage of Fidelity of Implementation's Core Component to Capture Variation in Treatment Effects**

Lisa Hall Foster, Sarah Oh, Carolyn Callahan and Amy Azano, University of Virginia

ABSTRACT

The inclusion of fidelity measures in effectiveness research is a relatively nascent concept. Well-developed and valid fidelity measures help explain variance in outcomes and reduce errors in evaluation. Moreover, fidelity is a critical link between the intervention and student outcomes. The purposes of this study were to determine the extent to which the curricular intervention designed for third-grade gifted students was implemented, how fidelity is related to student learning outcomes, how observed fidelity of implementation (FOI) related to self-reported FOI, and why teachers made adaptations to the curricular intervention.

Results indicated that teachers were able to implement the intervention with a moderate degree of fidelity, making substantially more modifications to program model, than to actual subject-matter content critical components. Multilevel analyses indicated that teachers' adherence to the curricular intervention has a significant positive association with student outcome measures after controlling for students' prior achievement for both units ($p < .05$). Results also indicated that teacher self-reported FOI could be a reliable alternative to costly direct observations when measures are clear and precise. Finally, among the factors influencing teachers' implementation, most of the variables were related to internal factors, specifically teacher-related factors and school capacity. The overall findings provide valuable information on program feasibility and refinement, professional development for implementers, the value of using teacher self-report of implementation, and the continued construction of an empirical basis for defining and measuring fidelity of implementation.

Keywords: fidelity of implementation (FOI), assessment and measurement of fidelity, curricular interventions, self-report, factors influencing FOI

INTRODUCTION

The nucleus of effective gifted services is a curriculum that challenges and enhances learning outcomes for gifted students. Yet, there is limited research documenting the impact of different curricular models or materials in gifted education (VanTassel-Baska 2003). Furthermore, the degree to which the interventions are implemented with fidelity is often overlooked or poorly measured and reported even though the inclusion of fidelity of implementation (FOI) measures

* The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305A060044, as administered by the Institute of Education Sciences, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Institute of Education Sciences or the U.S. Department of Education

in effectiveness studies have been in the literature for over a decade (Mowbray, Holder, Teague & Bybee 2003; O'Donnell 2008) and is receiving increased attention (U. S. Department of Education 2006).

When a program is developed and tested for effectiveness, the testing occurs in limited contexts. Fidelity assessments provide an evaluative systematic link between program implementation and outcomes attributable to the program's success or failure (Sanchez et al., 2007). Unless researchers have measured and assessed the fidelity with which the curricular intervention was implemented, neither the designer nor the implementer knows if the intervention model is truly effective or under what conditions it is or is not effective. Although it is widely known that there are variations in educational innovations and interventions when implemented (Blakely et al., 1987; Dusenbury, Brannigan, Falco, & Hansen, 2003), there is a point in variation that, when passed, would compromise the integrity of the program. Hence, fidelity measurements and assessments as part of empirical research, help to account for possible variance and its effects.

The goal of the present study is to understand: (a) the extent to which a gifted curricular intervention is implemented with fidelity, (b) how fidelity is related to student learning outcomes, (c) how teacher self-report of fidelity is associated with observed fidelity, and (d) teachers' reasons for making adaptations to a curricular intervention. This study responds to the call for increased understanding of fidelity, the construction of fidelity frameworks, effective data collection, and insights into how changes to different components of an intervention are related to fidelity. Additionally, the study expands our knowledge of how to facilitate higher levels of FOI.

The Clear Curriculum Model

The present study on fidelity was embedded within a larger study, "What Works in Gifted Education" (WWIGE). To develop empirical and descriptive understandings of model-based curricular units for gifted students, critical components from three well-known models in gifted education, namely Tomlinson's Differentiated Instruction Model, Renzulli and Reis' Schoolwide Enrichment Model (SEM), and Kaplan's Depth and Complexity Model, were integrated into a curricular model called the Clear Curriculum Model (CCM). Additionally, two language arts

units for third-grade gifted students were developed. CCM served as the theoretical and philosophical underpinning for two language arts/reading units. One unit focuses on poetry. The other unit focuses on expository, nonfiction text, and research skills (referred to as the "poetry" and "research" units, respectively).

The units emphasize conceptual thinking, real-world interdisciplinary inquiry, and problem-solving skills to encourage student production of authentic products. Throughout the units, differentiation is applied to address learner readiness, interest, and learning profiles. Principles of differentiation include ongoing assessment, appropriate challenge, and flexible grouping (Tomlinson, 1983; 2010). To build layers of challenge and meaning onto standards-based learning opportunities, elements of depth (big ideas, layers of the discipline, details, patterns, and rules) and complexity (multiple perspectives, interdisciplinary connections, unanswered questions, ethical issues, and changes over time) are incorporated into the curriculum (Kaplan, 2005). Consistent with the Schoolwide Enrichment Model (Renzulli & Reis, 1985; 2000), opportunities for students to work with tools and methods of practicing professionals and engage in "real-world" projects in an area of interest are integrated into the curriculum.

REVIEW OF THE LITERATURE

Fidelity of Implementation (FOI)

FOI measures are necessary to account for the moderating role that teachers play in the implementation of a curricular intervention. FOI, broadly defined, is the extent to which the delivery of an intervention, specifically its critical components, adheres to the program model and theory as originally developed (Century, Rudnick, & Freeman, 2008; Dusenbury et al., 2003; Lynch, 2007; Lynch & O'Donnell, 2005; Ruiz-Primo, 2006). There are multiple dimensions of fidelity (Dane & Schneider, 1998), including *adherence* (how closely the intervention is followed), *exposure* (the completeness of the delivery, e.g., number and length of sessions, or frequency), *quality of delivery* (how the intervention is implemented), *participant responsiveness* (the degree/level of participant's participation or engagement), and *program differentiation* (the presence or absence of the critical components that distinguish the intervention from the standard curriculum). For this study, FOI is operationalized as adherence, the degree to which teachers

implement lessons consistent with the intent of CCM.

Importance of Fidelity of Implementation Measures

The importance of monitoring implementation is undisputed in the fields of education, mental health, and psychology (Dusenbury et al., 2003), as the extent to which interventions are implemented varies greatly (Durlak & DuPre, 2008). Documentation and measurement of FOI, informs research as to whether unsuccessful outcomes reflect a poorly designed intervention or failure to implement an intervention as intended (Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005). Notably, studies conducted by Durlak and DuPre (2008) and Dusenbury et al. (2003) have already found a positive relation between higher levels of fidelity and intervention outcomes. Nonetheless, few education effectiveness studies, especially those related to gifted education, have measured or reported fidelity, much less used it to adjust for or interpret outcomes measures (O'Donnell, 2008). The inclusion of well-developed and valid fidelity measures helps explain variance in outcomes and reduces error in evaluation, promoting internal and external validity in outcome research (Dusenbury et al., 2003; Greenberg, Domitrovich, Graczyk & Zing, 2005; Mowbray et al., 2003; O'Donnell, 2008). The cost of not including fidelity measures is the implementation of ineffective programs or possible rejection of effective programs (Moncher & Prinz, 1991).

Measuring Fidelity of Implementation

A variety of methods and combinations of methods, both direct and indirect, can be used for measuring fidelity (Dusenbury et al., 2003), including direct observations, interviews, collection of permanent products, and self-report. Although the combination of direct observations and in-person interviews provide the most rigorous and complete measurement of FOI (Fullan & Pomfret, 1977), they require extensive financial, personnel, and time resources. Furthermore, to get a fuller measure, multiple observations and interviews are recommended (Fullan, 1983; Resnicow et al., 1998).

Self-report via surveys, questionnaires, logs, or checklists are an inexpensive and efficient measure for large-scale studies, providing a relatively accurate picture of classroom practice (Paulhus & Vaira, 2005; Ross, McDougall, & Hogaboam-Gray, 2003; Smithson & Ritz, 1994; Wobble's, Brekelmons, & Homomers, 1992). Studies examining the accuracy of teacher self-report measures relate mixed results. Self-report can be influenced by possible cognitive

limitations of an individual (Haugtvedt & Wegener, 1994; Tversky & Kahnemon, 1974), as well as teacher beliefs, teacher efficacy, and social bias (Ross et al., 2003). However, training in prevention concepts, (which can include professional development to improve implementation), provide techniques to counter social bias, or promote belief and buy-in, and can address these limitations and enhance the accuracy of self-report (Hansen & McNeal, 1999; Lillehoj, Griffin, & Spoth, 2004). Even with the limitations, Carroll et al. (2007) propose that self-report can help evaluate the extent to which participants fully accept the responsibilities required by the intervention and perceive the intervention to be useful. Correlations between self-report and practice can be strengthened if surveys, logs, and checklists are very clear and specific (Burstein et al., 1995); attempt to measure features that can be reported well using the self-report methods; and have high predictive validity (Ross et al., 2003). This study contributes to the understanding of how direct and indirect measures can be used in coordination to reduce financial expenses without sacrificing collection of valuable data necessary to measure FOI.

Assessing Fidelity of Implementation

According to research conducted by the National Diffusion Network, over half of the organizations adopting innovations modified the intervention (Rogers, 2003). As such, the establishment of effective criteria to measure and assess FOI is necessary to understand how those changes impact outcomes. O'Donnell (2008) suggests a systematic approach by identifying possible indicators or critical components of a given model, collecting data to measure the indicators, and examining the indicators in terms of their reliability and validity. When developing the approach, it is important to consider the levels of specificity, the degree of acceptable deviation, and the amount of qualitative and empirical data required to determine the fidelity (Ruiz-Primo, 2006). Through a focus on identification of critical components in relation to achieved fidelity levels and the criteria for determining acceptable deviation from the written curriculum, we have, in this study, demonstrated how this process could provide a meaningful fidelity measure of lessons to the theoretical framework of an intervention.

Factors Influencing Fidelity of Implementation

A multitude of internal and external factors influence FOI. A summary of the factors is depicted in Figure 1. Internal factors influence fidelity within the school's boundaries. Teacher characteristics, such as a teacher's age, career stage, race, culture (Datnow & Castellano, 2000);

skill proficiency (Durlak & DuPre, 2008); perceptions and expectations of the benefits of the intervention (Durlak & DuPre, 2008); need for ownership (Datnow & Castellano, 2000); prioritization (Brunette et al., 2008); beliefs, practices and school culture (Datnow & Castellano, 2000); as well as experience level (Huang & Moon, 2008; Rivkin, Hanushek, & Kain, 2005) can influence fidelity. School factors, or the organizational capacity of the school, can significantly influence FOI. For the most part, organizational capacity is a result of administrative leadership. Effective leadership skills support effective and successful implementation (Kam et al., 2003; Kurki et al., 2006). Higher levels of FOI are evidenced in the implementation of schoolwide programs when administrators cultivate a receptive climate by modeling positive attitudes and beliefs about the intervention, encouraging shared decision making to participate, communicating effectively, providing support and feedback, participating in the training, and managing the implementation as intended (Brunette et al., 2008; Century et al., 2008; Durlak & DuPre, 2008; Kurki et al., 2006; Proctor, 2004).

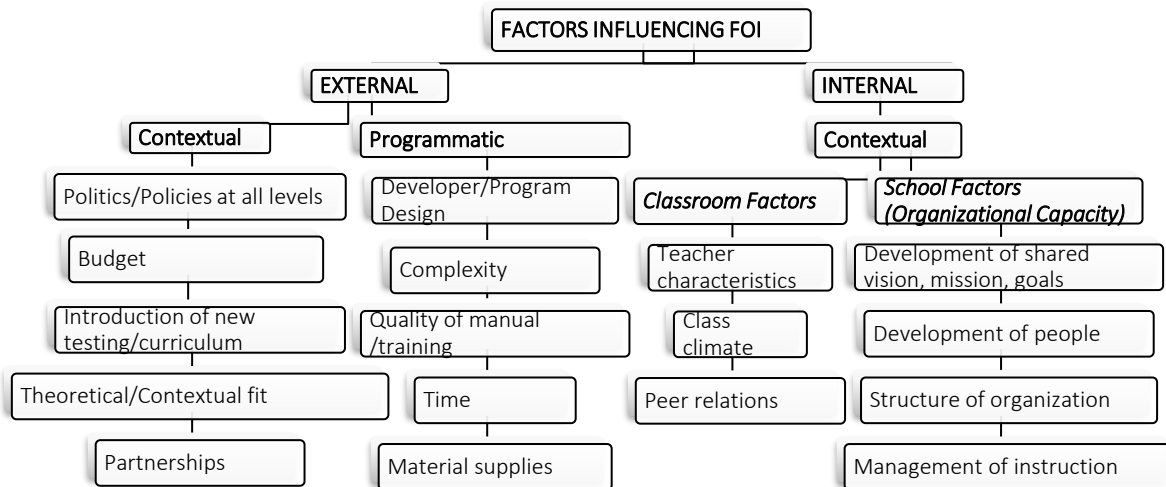


Figure 1 Factors influencing fidelity of implementation synthesized from this literature review

External factors relating to policies/politics, budgets, prioritization of mandates and monies, fit with district mission, partnerships and resources, multiple innovations, and community buy-in also influence implementation, but typically more indirectly (Brunette et al., 2008; Century et al., 2008; Datnow & Castellano, 2000; Everhart & Wandersman, 2000). Program factors, such as the design, complexity, and quality of the intervention, the amount of

time required for implementation, the supply of materials, the quality of training, and support from developers, can influence FOI (Ruiz-Primo, 2006). Given that most studies to date have focused on factors that influence FOI during the implementation of schoolwide or full grade-level interventions, this study contributes to the understanding of how and why individual teachers, in particular, non-core gifted teachers whose support levels may differ from core teachers, are influenced when deciding how to implement an intervention.

METHODS

Gaps in the literature illustrate the need for studies examining the development and inclusion of FOI measures in effectiveness studies, especially in the gifted education context. One goal of this study was the development of FOI measures for a classroom-based gifted curricular intervention to facilitate understanding of how self-report measures of adherence can be used effectively and efficiently as well as to examine the connection between FOI and student learning outcomes in this new context.

The researchers employed a mixed methods design to answer four research questions.

- To what extent is a curricular intervention implemented as intended in third-grade gifted classrooms?
- How is fidelity to the curricular intervention related to student outcomes?
- How is teacher self-report of fidelity related to observed fidelity?
- What are the reasons given by teachers for varying levels of fidelity?

Research Design and Participants

In the WWIGE study, a cluster-randomized experimental design was the basis for randomly assigning classrooms to treatment or comparison conditions. Teachers and students in pull-out or self-contained classrooms specifically designated for gifted students were recruited through national advertisement at the state and district level. Interested teachers were informed that participation was contingent on compliance with random assignment. After recruitment, classrooms were randomly assigned to experimental or comparison conditions with students clustered within classrooms. In cases where teachers taught at multiple sites, or there were multiple teachers in a school, those teachers were assigned to the same condition to avoid possible contamination of treatment effect from a teacher inadvertently using the curriculum in

comparison classrooms.

A concurrent-embedded mixed methods design guided this study with qualitative data used to inform quantitative results (Creswell & Plano Clark, 2007). Quantitative and qualitative data were collected and analyzed during various stages of the study to determine fidelity to the intervention. Observations and interviews were used to determine adherence to CCM. The adherence scores were then divided into typologies to provide a framework for the qualitative analysis of teachers' reasons for making adaptations, which in turn, informed the development of Teacher Fidelity Profiles which can be used to facilitate higher levels of FOI.

The sample for the current study included 30 teachers assigned to treatment schools in eight states (Colorado, Florida, Georgia, Indiana, Kentucky, South Carolina, Texas, and Virginia). Participants selected for the first round of observations reflected the overall composition of the population of the larger study. Eighteen participants were new to the study; 12 participated in the study the previous year. All the teachers were females teaching third-grade gifted students in pull-out or self-contained settings. Twenty-eight of the participants self-identified as White, one as Hispanic, and one as American Indian.

Procedures

In the fall of 2010, teachers in the treatment condition were given two curricular units to implement. Teachers could decide the order and scheduling of implementation of the units to be completed by the end of the school year. In addition to the written units, all materials needed to implement the units were included in the initial box of materials (i.e., pencils, post-it notes, index cards, pre-made copies of all handouts and assessments, notebook paper, binder, folders, return labels, and shipping envelopes). The WWIGE research team developed a webinar for teachers explaining the purpose of the study, the program model, the layout of the curriculum manual, and the use of materials. In addition, teachers were provided continuous access to instructional support via phone calls, emails, the website, and additional webinars.

Data Sources

Fidelity Data Sources

There were four fidelity data sources: observations, teacher logs, interviews, and school/teacher profiles. A member of the research staff conducted a 30 to 60 minutes observation of each

teacher. Observations were recorded on an observation guide, a semi-structured tool developed to parallel the critical components in the poetry and research units. The observer indicated whether each lesson component was implemented, modified, or omitted. In addition, observers were asked to record observational or analytical comments, descriptions of any modifications or deletions, and teacher and student actions or behaviors. Observers received training on the use of the guide. Reliability, assessed using a percent agreement obtained through a double-coding process, exceeded 80%.

Following each observation, the observer asked the teacher to complete the section of a teacher log that mirrored the observation guide. In the log, teachers reported how they perceived their implementation of each lesson using a checklist of items. The use of a checklist was predicated on the assumption that a teacher would be more likely to fill out the log and return it at the end of the unit if the form were easy to complete. Additionally, the form included space to provide descriptions of any modifications, omissions, or additions and the rationale behind the adaptation, as well as any additional information.

After the teacher completed the teacher log, the observer conducted a 15 to 30-minute interview that was recorded and later transcribed. Transcriptions were returned to the respective teacher for review and corrections. Data collected during the interview were guided by a semi-structured, open-ended interview protocol. The questions were designed to elicit information about the school's gifted program, readiness levels of students, class environment, experience or knowledge of poetry/research, the rationale for making or not making modifications or omissions, and the factors that influenced the teacher's decision to modify the curriculum.

Lastly, school and teacher characteristics (e.g., teaching experience, gifted teacher experience, educational attainment, and gifted certification/endorsement) were available from a survey administered prior to implementation of the intervention.

Student Data Sources

Student assessment data included both pre- and post-assessments. Student performance on the Iowa Tests of Basic Skills (ITBS) Survey Battery Reading subtest (Level 9, Form A) was administered at the beginning of the school year to ensure equity of achievement levels before the intervention. The Reading Comprehension subtest is designed to measure how students

derive meaning from reading passages representing narrative, poetry, and nonfiction material from science and social studies (Hoover et al., 2003). The internal consistency reliability estimate of the Reading subtest for the sample participants was .762. For the student outcome measurement, two standards-referenced post-tests with 35 items each were constructed by the research team and piloted at treatment sites in different states to determine the appropriate level of difficulty and discrimination for each item. Cronbach's alpha and Spearman-Brown reliability estimates ranged from .72 to .79 across the two tests.

Data Analysis

Level of Fidelity of Implementation

An extensive analytic approach over multiple phases was employed to determine the level of fidelity of implementation of model-based curricular units and association of fidelity with student learning outcomes.

First, the Design Principles and Content Critical Components of CCM were identified to facilitate a complete understanding of the program model. The Design Principles were representative of the best practices in gifted education that CCM embodied. The Content Critical Components entailed content requisite in the successful education of poetry and research/informational text. Because CCM is derived from three different models in gifted education, codes were cross-referenced between the models for analysis and interpretation. For example, grouping arrangement/flexible grouping occurs in both the Differentiation of Instruction Model (Tomlinson, 2010) and the Schoolwide Enrichment Model (Renzulli & Reis, 1985; 2000). The cross-referencing provided a visual for where the integration and synthesis of the three models occurred, and a means for accounting for a single concept that was evidenced in more than one model. When descriptive statistics were run on codes, cross-referenced codes were accounted for only once. Six lessons (three from each unit) were double-coded by two researchers to obtain inter-rater reliability on the coding of design principles and critical components.

Any discrepancies that occurred were resolved by discussing and reaching consensus as to the components' relevance in CCM.

In the second phase, data were merged (Caracelli & Green, 1993) and steps were taken

to ensure reliable scoring of adherence. Using the observation guide and interview transcript, four observations were initially scored at a level that was reflective of the intent of CCM, recording on a summary sheet whether each component was implemented as designed, modified, or omitted, using the Design Principles to facilitate the process.

For each observation, components implemented as designed or modified/omitted in accord with CCM framework were assigned a point value of one. Modifications/omissions not in accord were assigned a zero value. The values were summed and divided by the total number of components in the lesson to obtain an observed fidelity percentage. To ensure that scoring accurately reflected an in-depth knowledge of CCM and the nuances of change considered to be in accord with the program model, inter-rater reliability using a Kappa statistic was computed on the four sets of observations. Kappa exceeded 80% on each. A set of Guiding Principles for determining the acceptability of modifications or omissions was developed to guide decisions of acceptability of adaptations/omissions. The development of the Guiding Principles necessitated an iterative process. In the final phase, researchers scored the remaining observations using the Guiding Principles.

Analysis of Relationship of FOI to Student Outcomes

After determining the levels of fidelity, we employed another purposive sampling strategy, maximum variation sampling, to understand a wide range of teacher fidelity levels and their impact on student outcomes (Creswell, 2007; Miles & Huberman, 1994). Maximum variation sampling is appropriate for sequential methodological approaches where conclusions drawn at one phase of a study (varying level of fidelity of implementation) can then be used to inform further direction of the study (Creswell, 2007). This sampling strategy resulted in a reduced sample size of 15 teachers, 10 of whom were categorized as high fidelity and five as low fidelity, for subsequent quantitative analysis. This methodological approach allowed more precise exploration into the relationship between the level of teacher fidelity and student outcomes in a manner consistent with the purposes of maximum variation sampling.

Multilevel analyses of post-test scores for both units allowed the nested nature of the data set to be considered and prevented issues with aggregation bias and the misinterpretation of standard errors in the analysis (Raudenbush & Bryk, 2002). All analyses used a statistical

package Mplus 6 that allowed straightforward and exact parameter estimation considering a small cluster size for the multilevel analyses in the current study (Muthen, 2010).

Prior to data analysis, intra-class correlation (ICC) and design effect (DE) statistics were calculated to decide whether multilevel modeling was necessary. There was significant between-classroom variance: The ICC ranged from 0.18 to 0.39 for both units. In addition, the DEs, which ranged from 2.89 to 5.06, also supported the need for multilevel modeling for outcome measurement for both units.

To examine the relationship between teachers' level of adherence to the curricular intervention and student learning outcome measures, multilevel models for each unit were generated. The Level 1 model contained students' ITBS scores. The Level 2 model included teacher's adherence to the curricular intervention. Based on the qualitative data analysis with maximum variation sampling (Creswell, 2007; Miles & Huberman, 1994), teachers' adherence to the curriculum was coded as 0 = *low-fidelity group* and 1 = *high-fidelity group*. The pre-test ITBS scores were grand mean centered as a proxy for previous unaccounted influences. The level 1, level 2 and combined models are presented below.

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} (\text{ITBS}_{ij} - \overline{\text{ITBS}}..) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} (\text{Fidelity}) + \mu_{0j}$$

$$\text{Level 2: } \beta_{1j} = \gamma_{10}$$

$$\text{Combined: } Y_{ij} = \gamma_{00} + \gamma_{01} (\text{Fidelity}) + \gamma_{10} (\text{ITBS}_{ij} - \overline{\text{ITBS}}..) + r_{ij} + \mu_{0j}$$

Self-reported Fidelity and Observed Fidelity

The teacher self-reported fidelity was computed in the same manner as the observed fidelity percentage using the teacher's log, which mirrored the Observation Guide. A correlational analysis was conducted to test the relationship between teacher self-reported fidelity and observed fidelity. As the data were not normally distributed, a Spearman (ρ) correlation coefficient was calculated.

Typologies from observed fidelity scores were developed to understand the reasons given by teachers for making changes to the intervention in relation to their fidelity levels. Teachers

were grouped into one of three levels of fidelity: High (85-100), Moderate (70-84), or Low (69 and below). Although the division indicators are higher than suggested by Borelli et al. (2005), every component in each lesson reflected either a design principle or content critical component such that implementation, as written or intended, was important to achieve the desired outcome purported by CCM. Implementation below 70% would compromise the integrity of CCM. These three dimensions were the basis for within-group and between-group qualitative analyses to form teacher fidelity typologies. Researchers proceeded to analyze interview data and teacher logs inductively to understand teachers' rationale for making modifications. The inductive process encompassed reading, searching, coding, and analyzing the data; making assertions, generating confirming and disconfirming evidence; testing the assertions against the data; and returning to discrepant data (Creswell, 2007). Triangulated data methods, member checking, and an external audit all contribute to the study's plausibility.

RESULTS

Level of Fidelity to the CLEAR curriculum Model

Observed fidelity scores were reported as percentages. The score is the sum of the critical components implemented as intended or modified/omitted in accord with CCM's design principles. A summary of overall and categorical descriptive statistics is included in Table 1. The overall mean for observed fidelity was 83.27% ($SD = 20.39$), which fell into the range of moderate fidelity based on the pre-set fidelity levels. Eighteen teachers implemented a lesson with high fidelity, five with moderate fidelity, and seven with low fidelity. Five teachers implemented the lesson with 100% fidelity. Teachers with self-contained programs had higher fidelity (92%) than those in pull-out settings (79%). As the length of teaching time of each observation varied extensively between self-contained and pull-out settings, intervention developers should consider this matter with regard to design and anticipated adaptations. New teachers had a slightly higher mean (84.39%) than returning teachers (81.58%). Notably, the sample size was very small, and the standard deviations were large. The large standard deviations were due to high variability in the data and the outlier.

Table 1 Summary of Observed Fidelity Scores

Observed FOI Score		With Outlier			Without Outlier	
		N	Mean	Std. Deviation	Mean	Std. Deviation
Overall	Overall	30	83.27	20.39	85.28	17.47
Teacher Status	Returning Teachers	12	81.58	23.92	86.73	16.74
	New Teachers	18	84.39	18.32	84.39	18.32
Program Type	Self -contained Program	9	92.11	11.70	92.11	11.70
	Pull -out Program	21	79.48	22.31	82.20	18.97
Endorsement	Endorsed	10	77.5	23.91	83.13	16.13
	Not endorsed	20	86.15	18.37	86.15	18.37
Certification	Certified	10	83.00	20.00	83.00	20.00
	Not certified	20	83.00	21.00	86.32	16.70
Masters Degree	Degree	15	86.00	22.00	90.86	14.48
	No Degree	15	80.00	19.00	80.00	19.00
Gifted Masters	Gifted Masters	5	98.00	4.47	98.00	4.47
	No Gifted Masters	25	80.32	21.09	82.63	18.04
Fidelity Rating	High	18	97	6	97	6
	Moderate	5	78	4	78	4
	Low	7	52	15	57	10

Modifications and Omissions to the clear curriculum Model

In addition to determining whether the observed lessons were implemented with high, moderate, or low fidelity, a qualitative analysis was used to support the quantitative findings, highlighting where the changes to the curriculum occurred.

Design Principles and Content Critical Components

Teachers made 356 adaptations to the curriculum, with almost twice as many changes to research lessons than poetry lessons. Adaptations (modifications or omissions) to the CCM were more likely to occur to Design Principles (63%) as opposed to Content Critical Components (37%). The type of components changed, were similar; however, the number of changes to different categories varied. Over 80% of the changes to Design Principles occurred in nine of the 17 categories, with the majority of those nine occurring in five categories: (a) Exploration of details,

patterns, rules, ethics, and unanswered questions; (b) Procedural tasks; (c) Use of multiple modes of flexible grouping; (d) Exploration of generalizations and big ideas indicative of quality curriculum; and (e) Assignment of tasks based on student readiness, interest, and learning style (see Table 2). Adaptations to Content Critical Components shared by teachers in both units occurred with: (a) Big Ideas, (b) Procedural Tasks, (c) Reading skills and (d) Vocabulary.

Table 2 Overall Modifications to the Design Principles

Design Principle	Code	Number of Overall Changes (N= 255)	Percent of Overall Changes
Exploration of details, patterns, rules, ethics, and unanswered questions	D-DPR	31	12
Procedural Tasks	RP/PP	31	12
Use of multiple modes of flexible grouping	DI-FGMM/SEM-GA	26	10
Exploration of generalizations and big ideas indicative of quality curriculum	DI-QC/D-BI	26	10
Assignment of respectful tasks based on student readiness, interest, and learning style	DI-RT	24	9
Facilitation of materials and tools of the discipline	SEM-TM	23	9
Continuous assessment	DI-CA	16	6
Flexible grouping	DI-FG	15	6
Literary focus in reading comprehension	LF-RC	15	6
Facilitation of real-world applications (Authentic Products)	SEM-AP	12	5
Facilitation of multiple perspectives	C-MP	11	5
Student centered learning environment/facilitation of self-directed study	DI-SLE/SEM-FS	6	2
Facilitation of language of the discipline	D-FL	5	2
Literary Focus-Word study	LF-WS	5	2
Performance-based assessments	SEM-PBA	4	2
Multiple modes of continuous assessment	DI-CA-MM	3	1
Literary Focus- Writing	LF-W	2	1

Results by Fidelity Grouping

Analysis by fidelity grouping provides information for designers to support individual teachers

during the implementation phase and researchers during the refinement cycle (Greenberg et al., 2005). Teachers with high fidelity made more modifications (68%) than omissions (32%) to the curriculum. Moreover, 60% of these changes were in accord with *CCM* design. Teachers with high fidelity made most of their changes (52%) to areas involving: (a) Exploration of designs and patterns (19%), (b) Big ideas (19%), and (c) Multiple modes of flexible grouping (14%).

Teachers with moderate fidelity also made more modifications (72%) than omissions (28%); however, the modifications were not usually in accord with *CCM*. Over 60% of their changes were equally spread over six areas: (a) Tasks based on student interest and readiness (16%), (b) Use of tools and materials unique to a discipline (11%), (c) Flexible grouping (10%), (d) Multiple modes of grouping (10%), (e) Exploration of designs and patterns, and (f) Reading comprehension.

Teachers with low FOI made twice as many adaptations as the other teachers combined. Furthermore, the adaptations involved making over twice as many omissions (70%) as modifications (30%). Notably, out of the entire sample of 30, the eight teachers who had low fidelity made 46% of all the adaptations to design principles.

Association between Level of Adherence and Student Learning Outcomes

The multilevel analyses indicated that teachers' adherence to the curricular intervention has a significant association with student outcome measures after controlling for students' prior achievement for both units ($p < .05$). Students in high-adherence teachers' classrooms outperformed students in low-adherence teachers' classroom. Model fit tests corroborated the significant association between teachers' fidelity to the curriculum and student outcome for both units: $\Delta\chi^2_{df=1} = 7.72, p < .001$ for the poetry unit and $\Delta\chi^2_{df=1} = 4.44 p < .05$ for the research unit, Global effect size for each unit was calculated. Results indicated that a significant amount of the variance in student outcome (29% for the poetry unit and 20% for the research unit) could be explained by teachers' adherence to the curricular intervention (see Tables 3 and 4).

Table 3 Model Summaries for Poetry Unit

	Unconditional		ITBS only		ITBS+ Fidelity	
	One-way ANOVA					
Parameter	Parameter	SE	Parameter	SE	Parameter	SE
	Estimate		Estimate		Estimate	
Intercept (γ_{00})	24.10**	.66	24.54***	.545	22.60***	.547
ITBS (γ_{10})	-	-	.112***	.03	.111***	.003
Fidelity (γ_{01})	-	-	-	-	2.837*	.886
Residual (σ^2)	14.83**	3.26	11.721**	2.34	11.656**	2.29
Intercept (τ_{00})	5.134**	2.12	2.857	1.93	1.123	.915
Deviance Statistic	874.408		747.856		740.138	
Number of estimated parameters	3		4		5	

Table 4 Model Summaries for Research Unit

	Unconditional		ITBS only		ITBS+ Fidelity	
	One-way ANOVA					
Parameter	Parameter	SE	Parameter	SE	Parameter	SE
	Estimate		Estimate		Estimate	
Intercept (γ_{00})	25.114***	.52	25.49***	.422	24.16***	.728
ITBS (γ_{10})	-	-	.101**	.024	.098**	.024
Fidelity (γ_{01})	-	-	-	-	1.821**	.825
Residual (σ^2)	14.80**	2.36	12.85**	1.93	12.88	1.948
Intercept (τ_{00})	2.477**	.899	1.021	.803	.360**	.585
Deviance Statistic	916.732		822.518		818.074	
Number of estimated parameters	3		4		5	

Relationship between Teacher Self-reported and Observed Fidelity

Descriptive Statistics

A summary of the overall and categorical statistics for self-reported fidelity is given in Table 5. The overall mean for self-reported fidelity was 90.17 ($SD = 13.46$). A visual examination of the means did not suggest notable differences in any of the sub-categories. Large standard deviations were due possibly to the high variability in the self-reported fidelity, as there were no outliers. When comparing the self-reported scores to the observed scores by fidelity group, the likelihood of the teacher reporting the same rank as the observer fell on a continuum of "very likely" to "not very likely." The higher the fidelity rating, the more likely the teacher was to report the implementation of the curricular intervention in a manner consistent with the observer; the lower the fidelity, the less likely. The teachers in the high fidelity group averaged a 2.5 points difference from the observer scoring. Thirteen of the 18 teachers matched the observer's score exactly, while three were seven to 15 points higher and two were six to seven points lower. Thirteen of the teachers reported scores within five percent of the observed rating. Three teachers rated their adherence within a 10% differential. Two of the high fidelity teachers reported score differences of greater than 10%.

Table 5 Summary of Teacher Self-reported Fidelity Scores

	Self-reported FOI Score	N	Mean	Std. Deviation
Overall	Overall	30	90.17	13.46
Teacher Status	Returning Teachers	12	89.67	16.04
	New Teachers	18	90.5	11.94
Program Type	Self-contained Program	9	97.89	3.26
	Pull-out Program	21	86.86	14.84
Endorsement	Endorsed	10	88.1	16.27
	Not endorsed	20	91.20	12.16
Certification	Certified	10	88.5	14.86
	Not certified	20	91.00	13.03
Masters Degree	Degree	15	90.07	14.84
	No Degree	15	90.27	12.46
Gifted Masters	Gifted Masters	5	98.00	4.47
	No Gifted Masters	25	88.60	14.15
Fidelity Rating	High	18	97.85	4.66
	Moderate	5	86.4	11.29
	Low	7	73	13.96

Teachers in the moderate fidelity group averaged a 12.8 points difference from observed scores, with four of the five teachers reporting higher scores than observed data. One teacher reported her score with less than a 5% differential and one with a 10% differential. Three of the five had over 10% differential in reporting. Finally, the lower fidelity group averaged a 20.6 points difference. Six of the seven teachers over-reported, meaning they reported higher percentages of fidelity than observed data. One teacher reported in a manner that resulted in the same percentage score and one reported within a 5% differential. Five of the seven had greater than a 10% differential.

Correlational Analysis

After determining the self-reported fidelity scores, a Spearman's rho correlation was calculated because the data from the observed and self-reported fidelity scores were not normally distributed. The skewness and kurtosis values were not within acceptable limits (> 1.0) and highly negatively skewed. There was a strong positive correlation between the two variables, $r_s(28) = 0.820$, $n = 30$, $p < 0.000$. In checking for the explained variance, $r_p^2 = .6724$, meaning 67% of the variance in self-reported fidelity is explained by observed fidelity. While there is a high correlation, actual scores and recording of particular components were different –more so as teachers deviated from fidelity (see above). As a result, self-report can be considered an acceptable alternative to direct observations to determine relative fidelity to CCM.

Reasons for Making Modifications to the clear curriculum Model

A number of factors were identified by teachers as contributing to their decisions to alter delivery of the curriculum. The factors fell into two overarching categories: internal to the school or external. The internal factors include attributes of the teacher, students, class, and school/school administration. The external factors were representative of anything to do with the curricular intervention or factors outside of the school's physical boundaries such as state and federal mandates, partnerships with outside organizations, budgetary restraints, the adoption of multiple interventions, or the introduction of a new curriculum. Eighteen of the factors were contextual, and three were programmatic (see Table 6). All the factors were consistent with factors identified in prior literature as contributing to deviance from fidelity.

Journal of Academic Perspectives

Table 6 Factors Influencing Teachers' Decisions to Make Adaptations to the CCM
 Factors Influencing Teachers' Decisions to Make Adaptations to Intervention

Internal	External
Knowledge of what is best for class (T)	Politics/Policies (CX)
Knowledge of students' abilities/skills (T)	Assessments/New Curriculum running concurrently with intervention (CX)
Perceptions and expectations (T)	Curriculum Design (P)
Unintentional Omissions (T)	Curriculum Complexity (P)
Intentional Omissions (T)	Time to implement intervention (P)
Comfort level (T)	
Ownership (T)	
Prioritization of the intervention (T)	
Concurrence with program philosophy/fit (T)	
Prior knowledge/completion (S)	
Student interests (S)	
Climate (C)	
Size (C)	
Time constraints (O)	
Structuring of organization (schedules/class size/subjects) (O)	
Instructional management (O)	

Note. Internal factors were contextual and consisted of four categories: teachers (T), students (S), classroom (C), and organizational capacity (O). External factors were either contextual or programmatic: contextual (CX) at local, state or federal level and programmatic (P).

Although teachers voiced many reasons for making adaptations (i.e., ownership, comfort level, prioritization, class climate, size), there were six factors most often associated with making modifications or omissions that were common to teachers at all three fidelity levels: (a) teacher knows best, (b) teacher knowledge of student abilities/skills, (c) time constraints, (d) teacher perceptions and expectations, (e) unintentional omission, and (f) organization within the school/administration. A summary is provided in Table 7.

Table 7 Summary of Top Seven Reasons for Making Changes to CCM

Top Seven Reasons for Making Modifications/Omissions	Overall Tally (N=190)	Percent of Total Reasons Given
Teacher knows best*	39	21
Teacher knowledge of student ability, skills and prior knowledge*	38	20
Time constraints	21	11
Teacher perceptions, expectations, beliefs, and practices*	16	8
Unintentional Omission*	13	7
Organization of the school by administrator	12	6

*Teacher-related factors

Internal Factors

Of the top five reasons for making changes, four (accounting for 60% of all reasons) were teacher related. Changes made because the teacher felt she knew best resulted in acceptable and unacceptable changes or omissions to the intervention. Typically, these changes were positive when based on knowledge of student skill/ability or prior knowledge as opposed to perceptions and non-data based expectations. Decisions based on teacher perceptions or expectations were never acceptable. Teachers with high fidelity made more decisions based on reported knowledge of students based on assessment, whereas teachers with low fidelity based more of their decisions on undocumented perceptions of student knowledge or ability. Although teachers at all three fidelity levels indicated that time constraints influenced decisions regarding implementation, the moderate level teachers indicated that time constraints weighed heavily in their ability to remember what they needed to complete during a lesson. Teachers with high fidelity could navigate time constraints successfully. Further research is suggested to determine why and how. Although most of the reasons teachers gave for making adaptations were common to all three fidelity levels (low, moderate, and high), the link between teachers' reasons for making adaptations and the level of fidelity was more the appropriateness of the change or omissions

than the rationale. High fidelity teachers were more likely to make decisions as to what to implement based on their knowledge of best practices in conjunction with knowledge of student ability/skill or student prior knowledge adaptations. The adaptations also were more likely to be in accord with the CCM. For example, one teacher who implemented the lesson with high fidelity made a modification to meet her students' needs as they had exceptionally high ability levels. In accord with CCM's premise that work should be matched with students' abilities, this teacher changed the text selections in the research unit to college-level texts. She explained:

It was, you know, a little easy for some of them, so I stepped it up and so my discussion, or in whatever I was teaching, um, we expanded. So, I brought in ... a specialist in research, I brought her in, and she spent a good hour and a half with my kids, and we did the BIG6 process with how we are going to search. She did Internet safety. She actually showed us about 25 different databases, so I did use the Internet sources that were in here, but then she probably gave us another 15-20. And then, we did a college-level text. (T95, pgs. 1-2)

Teachers with moderate fidelity made modifications based on student knowledge and prior assessments; however, their fidelity scores were mainly impacted by unintentional omissions or additions. For example, one teacher commented, "...sometimes I get on a roll with what I'm doing. I think on this one I simply didn't look down at the plans I think a lot of the changes I made were unintentional (T94, p. 3).

Teachers with low fidelity based their decisions on generalized perceptions and expectations of student knowledge/ability. In response to a question about letting her students work together on a pre-assessment, the teacher with low fidelity responded:

I find that when I use a circle map that the students, um, share information and they also bounce off of each other, piggyback on to other ideas and that sort of thing, and we get more information and more excitement than just filling out a piece of paper ...you know it is based on their knowledge, but they, with gifted people, gifted children they need to be working together and they feed on each other and, I think it just broadens them and you get many more, much more, you get better ideas, um, from doing that. (T34, p. 2)

Research, however, does not support the teacher's assumption that just because a child is gifted that they need to work collaboratively to broaden them and get better ideas (Renzulli et al., 2002; Tannenbaum, 1983; Torrance, 1974). Moreover, the teacher missed the purpose of the activity, which was to provide individual formative assessment data to inform instruction for the

next lesson and baseline data for researchers.

Organizational factors (i.e., scheduling, use of time and space, how tasks are assigned, and routine operating procedures) were mentioned the second most frequently. The school's principal controls all of these factors, and the teacher had no control. In one school, the principal allotted only 30 minutes a day, two days a week, for grade-level gifted instruction. According to the teacher, she made modifications trying to recap and bridge the previous lesson with the current lesson because 30 minutes was never enough time for her to complete an entire lesson.

I really think I tried to follow the lesson pretty closely. We did have the time constraints because we just have a half-hour with these guys, but I really tried to build on what we had done previously. I tried to draw their attention back to the thinking maps that we had done for the brainstorming to kind of get them back into thinking about what exploration is. (T43, p. 5)

External Factors

External factors did not have as big an impact on implementation. School, district, state, and federal level policies and district-mandated implementation of concurrent interventions were two contextual factors mentioned by teachers as influencing implementation. Programmatic factors, such as curriculum design and complexity, and the time needed to implement the curriculum were mentioned by teachers, but rarely.

Summary

There are several continua of the relationship between teachers at differing fidelity levels (see Figure 2). Teachers with high fidelity were more likely to make very few modifications, with those made, being in accord with the program model and based on actual knowledge of best practices in the classroom and student knowledge/ability. Teachers with low fidelity were likely to make multiple adaptations, with most of them not being in accord with the program model and based on generalized assumptions about student knowledge/ability. Moderate fidelity teachers fell in the middle on all three continua, but many of these teachers often unintentionally omitted items when pressured by time constraints causing them to fall in the moderate fidelity level, to begin with.

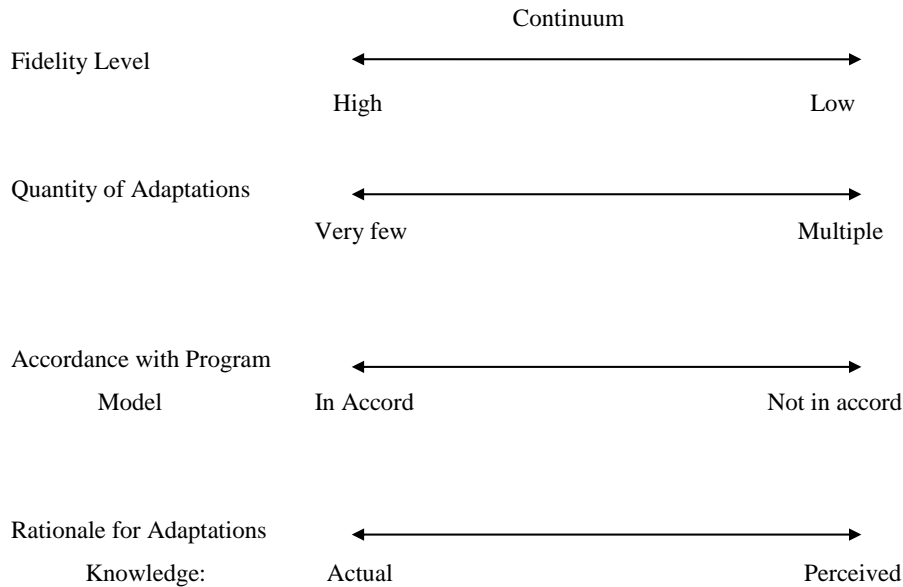


Figure 2 Fidelity continua for adaptations

Overall, the rationale given by teachers for making adaptations revealed that actual and perceived knowledge had a large impact on the fidelity level. Additionally, although time was a concern for all three fidelity levels, some teachers could work through the issue, and some were not. From the findings, Teacher Fidelity Profiles were developed (see [Table 8](#) in the Appendix).

DISCUSSION

In conjunction with the study of the extent to which teachers implemented experimental curricular intended, data which allowed examination of the degree to which fidelity was associated with student outcomes, how teacher self-report of fidelity related to observed fidelity, and why teachers made modifications to the lessons were collected. Although all teachers participating in the WWIGE study were given the same materials, professional development, and continuous support to encourage higher levels of fidelity and a better understanding of the program model, not all teachers implemented the units with the same level of fidelity. Overall, teachers implemented the lessons with moderate fidelity, less than two percentage points from implementing with high fidelity. Teachers modified or omitted practically the same design principles and critical components, but the extent to which the modifications or omissions were in accord with the program model varied greatly between fidelity levels.

Without the development and use of the Design Principles, as well as the identification of critical components, the measurement of the extent of fidelity of implementation and acceptability of modifications would not have been possible. This necessary step afforded researchers a second measure of fidelity to the curriculum's theoretical framework, in addition to an opportunity to make improvements in professional development of teachers and refinements to the curriculum itself (Greenberg et al., 2005; Nelson, Cordray, Hulleman, Darrow, & Sommer, 2010; Resnick et al., 2005; Ruiz-Primo, 2006). This step adds a new layer to the foundation of fidelity studies in curricular interventions not implemented schoolwide and imparts information necessary for scale-up. The identification of the Design Principles and Critical Components promotes creation of new units for gifted education (different content or different grade levels) by others based on the theoretical concepts and the patterns for integrating the components with the theory. With a deeper understanding of the theoretical framework of the curriculum, teachers may be better equipped to make appropriate adaptations when necessary (Greenberg et al., 2005; Resnick et al., 2005; Ruiz-Primo, 2006). Moreover, the use of fidelity measures facilitates replication of findings (Mowbray et al., 2003).

Not only did teachers' level of implementation fidelity differ, so did their accuracy of self-reporting fidelity. Overall, the study findings demonstrated a statistically significant positive relationship between teacher and observer implementation ratings. These findings are consistent with prior literature (Lillehoj, Griffin, & Spoth, 2004; Smithson & Pritz, 1994; Wubbels, Brekelmons, & Hoomayers, 1992) and are significant in light of oppositional reports on the limitations of self-report (Ross et al., 2003). Regardless of their observed fidelity level, over half of the teachers (16 of the 30) could report their scores within a five percent or less differential from the observed fidelity score. These findings suggest that the desired minimum of two to three observations to assess fidelity could be reduced with the incorporation of clear, concise self-reporting measures, thus reducing the financial and personnel resources requisite of direct observations (Fullan, 1983; Resnicow et al., 1998).

There are several factors to consider; however, in interpreting this positive correlation. The first factor is the time frame given teachers to report their implementation. In the larger WWIGE study, teachers were asked to complete the measure as they implemented each lesson

and return the teacher log with the other data upon completion of a full unit. With the observer asking the teacher to complete the log immediately following the observed lesson, some teachers may have felt rushed to fill out the log while others may have filled out the log more effectively because the researcher was present.

The second factor is the discovery of the continuum upon which teachers fell in regard to self-reporting. Two plausible explanations for varying accuracy levels in self-reporting are teachers' understanding of the intervention model and their skill sets. Teachers with high fidelity, those who made modifications in accord with the model, appeared to have more understanding of the model, and not surprisingly, were more accurate in reporting fidelity—likely because of their understanding. Although the accuracy of reporting varied, the rankings were correlated, allowing for the assessment of relative fidelity. Teachers start an intervention with varying skill sets. To address the conceptual understanding of an intervention's framework and thus possibly a teacher's fidelity of implementation, Greenberg et al. (2005) recommend in-depth training in the program model theory during the pre-implementation phase as teachers start interventions with varying skill sets.

Although the teachers' reasons for making changes to the lessons had commonalities, there were differences that suggest further research. Several factors contributed to the decisions teachers made to modify the curriculum. Most of these factors were teacher-related; others were attributed to students, the class, school administration, and the program intervention. Teachers' reasons for making adaptations have implications for researchers studying fidelity during pre-adoption and delivery phases of implementation in that knowledge of which factors influence teachers' implementation practices provide information to address the factors prior to and during implementation in an effort to reduce barriers to implementation and facilitate higher levels of fidelity (Greenberg et al., 2005). Awareness of these factors affords administrators responsible for implementation of curricular interventions the ability to address school-level factors when possible.

LIMITATIONS

There were several limitations to this study. Only a snapshot of fidelity was obtained through single observations and interviews. Also, the observations were of varying lessons and lengths

of time. In creating an album of snapshots, however, the researchers were able to provide fidelity profiles contributing to the scant empirical research on FOI. The small sample size limited analyses between high and low fidelity groups and analysis of the impact of critical components on outcomes.

CONCLUSION

The findings and implications from this study on fidelity are significant to the field of education and gifted education, as well for the construction of an empirical base for assessment and measurement of fidelity. This study demonstrated that it is not enough to know the level of fidelity with which a teacher implements an intervention. The development of nuanced adherence measures, which considered the acceptability of adaptations to the program model provides an example for the design of a more robust adherence measure, especially when resources for measuring FOI are limited. Furthermore, with the development of teacher fidelity profiles, which illuminate characteristics typical of teachers at varying fidelity levels, aggressive steps to facilitate higher fidelity levels and circumvent actions and choices indicative of lower levels prior to and during implementation can be taken. Evidence linking FOI with student learning outcomes, in addition to the development of a nuanced measure of adherence, expands not only the empirical base of fidelity research but also, the limited empirical base for evaluations of gifted education teachers and the curricula they implement.

REFERENCES

- Blakely, C., J. Mayer, R. Gottschalk, N. Schmitt, W. Davidson, D. Roitman, et al. "The Fidelity-Adaptation Debate: Implications for the Implementation of Public Sector Social Programs," *American Journal of Community Psychology* 15, no. 3 (1987): 253-268. doi:10.1007/BF00922697.
- Borrelli, B., D. Sepinwall, A.J. Bellg, R. Breger, R., C. DeFrancesco, D. Sharp, D. Orwig. "A New Tool to Assess Treatment Fidelity and Evaluation of Treatment Fidelity Across 10 Years of Health Behavior Research," *Journal of Consulting and Clinical Psychology* 73, no. 5 (2005): 852-860. doi:10.1037/0022-006X.73.5.852.
- Brunette, M., D. Asher, R. Whitley, W. Lutz, B. Wieder, A. Jones, and G. McHugo. "Implementation of Integrated Dual Disorders Treatment: A Qualitative Analysis of Facilitators and Barriers," *Psychiatric Services* 59, no. 9 (2008): 989-995. doi: 10.1176/appi.ps.59.9.989.

- Burstein, L., L. McDonnell, J. Van Winkle, T. Ormseth, J. Microchip, and G. Glutton. *Validating National Curriculum Instruction*. Santa Monica, CA: RAND, 1995.
- Caracelli, V., and J. Greene. "Data Analysis Strategies for Mixed-Method Evaluation Designs," *Educational Evaluation and Policy Analysis* 15, no. 2 (1993): 195-207. doi: 10.3102/01623737015002195.
- Carroll, C., P. Patterson, S. Wood, A. Booth, J. Rick, and S. Balain. "A Conceptual Framework for Implementation Fidelity," *Implementation Science* 2, no. 40 (2007): 1-9. doi:10.1186/1748-5908-2-40.
- Century, J., C. Freeman, and M. Rudnick. "A Framework for Measuring and Accumulating Knowledge About Fidelity of Implementation (FOI) of Science Instructional Materials," Presentation at the National Association for Research in Science Teaching, Baltimore, MD. (March 2008) Retrieved from <http://cemse.uchicago.edu/node/37>.
- Creswell, J. W. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Thousand Oaks, CA: Sage, 2007.
- Creswell, J. W., and V. L. Plano Clark. *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage, 2007.
- Dane, A. V., and B. H. Schneider. "Program Integrity in Primary and Early Secondary Prevention: Are Implementation Effects out of Control," *Clinical Psychology* 18 (1998): 23-45. doi:10.1016/S0272-7358(97)00043-3.
- Datnow, A., and M. Castellano. "Teachers' Responses to Success for All: How Beliefs, Experiences, and Adaptations Shape Implementation," *American Educational Research Journal* 37, no. 3 (2000): 775-799. doi:10.3102/00028312037003775.
- Durlak, J., and E. DuPre. "Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation," *American Journal of Community Psychology* 41 (2008): 327-350. doi:10.1007/s10464-008-9165-0.
- Dusenbury, L., R. Brannigan, M. Falco, and W. B. Hansen. "A Review of Research on Fidelity of Implementation: Implications for Drug Abuse Prevention in School Settings," *Health Education Research* 18, no. 2 (2003): 237-256. doi:10.1093/her/18.2.237.
- Dusenbury, L., R. Brannigan, W. B. Hansen, J. Walsh, and M. Falco. "Quality of Implementation: Developing Measures Crucial to Understanding the Diffusion of Preventive Interventions," *Health Education Research* 20, no. 3 (2005) 308-313. doi:10.1093/her/cyg134.
- Everhart, K., and A. Wandersman. "Applying Comprehensive Quality Programming and Empowerment Evaluation to Reduce Implementation Barriers," *Journal of Educational and Psychological Consultation* 11, no. 2 (2000): 177-191. Retrieved from <http://www.informaworld.com/smpp/title~content=t775653646~db=all>.
- Fullan, M. "Evaluating Program Implementation: What Can Be Learned from Follow Through." *Curriculum Inquiry*, 13, (1983): 215-227.

- Fullan, M., and A. Pomfret. "Research on Curriculum and Instruction Implementation," *Review of Educational Research* 47, no. 1 (1977): 335-397. Retrieved from <http://www.aera.net/publications/?id=319>.
- Greenberg, M. T., C. E. Domitrovich, P. A. Graczyk, and J. E. Zins. *The Study of Implementation in School-Based Preventative Interventions: Theory, Research, and Practice*, Volume 3. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration, 2005.
- Hansen, W., and R. McNeal. "Drug Education Practice: Results of an Observational Study," *Health Education Research*, 14, (1999): 85-97. doi:10.1093/her/14.1.85.
- Haugtvedt, C., and D. Wegener. "Messenger Order Effects in Persuasion: An Attitude Strength Perspective," *Journal of Marketing Research*, 17, (1994): 294-306. Retrieved from <http://www.marketingpower.com/AboutAMA/Pages/AMA%20Publications/AMA%20Journals/Journal%20of%20Marketing%20research/JournalofMarketingresearch.aspx>.
- Hoover, H. D., S. B. Dunbars, D. A. Frisbie, K. R. Oberley, V. L. Ordman, R. J. Naylor, G. P. Shannon. *The Iowa Tests: Guide to Research and Development*. Itasca, IL: Riverside, 2003.
- Huang, F., and T. Moon. "Is Experience the Best Teacher? A Multilevel Analysis of Teacher Characteristics and Student Achievement in Low Performing School," *Education Assessment Evaluation Accountability* 21, no. 3 (2009): 209-234. doi:10.1007/s11092-009-9074-2.
- Kam, C., M. Greenberg, and C. Walls. "Examining the Role of Implementation Quality in School-Based Prevention Using Paths Curriculum," *Prevention Science* 4, no. 1 (2003): 55-63. Retrieved from <http://www.preventionscience.org/prevscience.php>.
- Kaplan, S. "Layering Differential Curricula for Gifted and Talented." In *Methods and Materials for Teaching Gifted Students*, edited by F. Karnes and S. Bean, 107-132. Waco, TX: Prufrock Press, 2005.
- Kurki, A., A. Boyle, and D. Aladjem. "Implementation: Measuring and Explaining the Fidelity of CSR Implementation," *Journal of Education for Students Placed at Risk* 11 (2006): 255-277. doi:10.1207/s10824669espr110304_3.
- Lillehoj, C., K. Griffin, and R. Spoth. "Program Provider and Observer Ratings of School-Based Preventive Intervention Implementation: Agreement and Relation to Youth Outcomes," *Health Education and Behavior* 31, no. 2 (2004): 242-257. doi:10.1177/1090198103260514.
- Lynch, S. A. *Model for Fidelity of Implementation in a Study of a Science Curriculum Unit: Evaluation Based on Program Theory*. Paper presented at the annual meeting of the American Educational Research Association in Chicago, IL. (2007). Retrieved from <http://www.gwu.edu/~scale-up/>.
- Lynch, S., and C. O'Donnell. The Evolving Definition, Measurement and Conceptualization of Fidelity of Implementation Scale-Up of Highly Rated Science Curriculum Units in

- Diverse Middle Schools. Paper presented at the symposium on Fidelity of Implementation at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada. (2005, April) Retrieved from <http://www.gwu.edu/~scale-up/>.
- Muthén, B. Bayesian Analysis in Mplus: A Brief Introduction. Technical Report. Version 3 Los Angeles, CA: Muthén & Muthén, 2010.
- Miles, M., and A. M. Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd edition. Thousand Oaks, CA: Sage Publications, 1994.
- Moncher, F., and R. Prinz. "Treatment Fidelity in Outcome Studies," *Clinical Psychology Review* 11, no. 3 (1991): 247-266. doi:10.1016/0272-7358(91)90103-2.
- Mowbray, C. T., M. C. Holter, G. B. Teague, and D. Bybee. "Fidelity Criteria: Development, Measurement, and Validation," *American Journal of Evaluation* 24, no. 3 (2003): 315-340. doi:10.1177/109821400302400303.
- Nelson, M. C., D. S. Cordray, C. S. Hulleman, C. L. Darrow, and E. Sommer. *A Procedure for Assessing Fidelity of Implementation in Experiments Testing Educational Interventions*. Paper presented at the 2010 Society for Research on Educational Effectiveness conference (March 2010), Washington, DC.
- O'Donnell, C. L. "Defining, Conceptualizing, and Measuring Fidelity of Implementation and its Relationship to Outcomes in K-12 Curriculum Intervention Research." *Review of Educational Research* 78, no. 1 (2008): 33-84. doi:10.3102/0034654307313793.
- Paulhus, D. and S. Vazire. "The Self-Report Method." In *Handbook of Research Methods in Personality*, edited by R. Robbins, C. Fraley, and R. Krueger, 224-239. New York: Guilford Press, 2007.
- Proctor, E. "Leverage Points for the Implementation of Evidence-Based Practice," *Brief Treatment and Crisis Intervention* 4, no. 3 (2004): 227-242. Retrieved from <http://www.clockss.org/clockss/Home>.
- Raudenbush, S. W., and A. S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition. London: Sage, 2002.
- Resnick, B., A. Bellg, B. Borrelli, C. DeFrancesco, R. Breger, J. Hecht, D. Sharp, S. Czajkowski. "Examples of Implementation and Evaluation of Treatment Fidelity in the BCC Studies: Where We are and Where We Need to Go." *Annals of Behavioral Medicine* 29 (2005): 46-54. doi:10.1207/s15324796abm2902s_8.
- Resnicow, K., M. Davis, M. Smith, A. Lazarus-Yaroch, T. Baranowski, and J. Baranowski. "How Best to Measure Implementation of School Health Curricula: A Comparison of Three Measures," *Health Education Research* 13, no. 2 (1998): 239-250. doi:10.1093/her/13.2.239.
- Renzulli, J. S. and S. Reis. (1985). *The Schoolwide Enrichment Model: A Comprehensive Plan for Educational Excellence*. Mansfield Center, CT: Creative Learning Press.

- Renzulli, J. S. and S. Reis. "The Schoolwide Enrichment Model." In *The International Handbook of Giftedness and Talent*, 2nd edition, edited by K. Heller, F. Mönchs, R. Sternberg, and R. Subotnik, 367-382. Oxford, England: Elsevier, 2000.
- Rivkin, S., E. Hanushek, J. Kain. "Teachers, Schools, and Academic Achievement," *Econometrica* 73, no. 2 (2005): 417-458. doi:10.1111/j.1468-0262.2005.00584.x.
- Rogers, E. M. *Diffusion of Innovations*, 5th edition. New York: Free Press, 2003.
- Ross, J., D. McDougall, A. Hogaboam-Gray, and A. LeSage. "A Survey Measuring Elementary Teachers' Implementation of Standards-Based Mathematics Teaching," *Journal for Research in Mathematics*, 34, no. 4 (2003): 344-363. Retrieved from <http://www.nctm.org/publications/jrme.aspx>
- Ruiz-Primo, M. (2006). *A Multi-Method and Multi-Source Approach for Studying Fidelity of Implementation*. CSE Report 677 for the National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles, CA. doi:10_1_1_119_505/rep=rep1type=pdf
- Sanchez, V., A. Steckler, P. Nitirat, H. Hallfors, H. Cho, and P. Brodish. (2007). "Fidelity of Implementation in a Treatment Effectiveness Trial of Reconnecting Youth," *Health Education Research* 22, no. 1 (2007): 95-107. doi:10.1093/her/cy1052
- Smithson, J. and A. Pritz. "Measuring Classroom Practice: Lessons Learned from the Efforts to Describe the Enacted Curriculum," *The Reform Upclose Study*. Madison, WI: Consortium for Policy Research in Education, 1994. Retrieved from <http://www.eric.ed.gov>.
- Tannenbaum, A. (1983). *Gifted children: Psychological and educational perspectives*. New York: Macmillan.
- Tomlinson, C. A. *How to Differentiate Instruction in Mixed-Ability Classrooms*, 2nd edition. Upper Saddle River, NJ: Pearson, 2010.
- Torrance, E. *Torrance Tests of Creative Thinking*. Bensenville, IL: Scholastic Testing Service, 1974.
- Tversky, A. and D. Kahnemon. "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185, no. 4157 (1974): 1124-1130. Retrieved from <http://www.sciencemag.org>.
- United States Department of Education. *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User-Friendly Guide*. Washington, DC: Institute of Education Sciences, 2003. Retrieved from <http://www.ed.gov/print/rschstat/research/pubs/rigorousvid/guide.html>.
- VanTassel-Baska, J., L. Avery, J. Struck, A. Feng, B. Bracken, D. Drummond, and T. Stambaugh. *The William and Mary Classroom Observation Scales Revised*. Williamsburg, VA: The College of William and Mary School of Education, Center for Gifted Education, 2003.

Wubblels, T., M. Brekelmons, and H. Hoomayers. "Do Teacher Ideals Distort the Self-Report of their Interpersonal Behaviors?" *Teaching and Teacher Education* 8, no. 1 (1992): 47-58. doi: 10.1016/j.ijer.2006.03.003.

APPENDIX

Table 8 Profiles of High, Moderate, and Low Fidelity Teachers

Profile of a High Fidelity Teacher	Profile of a Moderate Fidelity Teacher	Profile of a Low Fidelity Teacher
<p style="text-align: center;">STATISTICS</p> <p>Range of fidelity scores: 85-100</p> <p>Average Observed FOI score: 97</p> <p>Average Self-reported FOI scores: 98</p> <p style="text-align: center;">Most likely to make modifications: (in order of percent occurrences) to the following Design Principles</p> <ol style="list-style-type: none"> 1. Depth: Exploration of details, patterns, and rules 2. Big Ideas/Quality Curriculum 3. Flexible groups/Multiple modes of grouping 4. Facilitation of tools and materials of the discipline 5. Facilitation of multiple perspective <p style="text-align: center;">and a few Content Critical Components</p> <ol style="list-style-type: none"> 1. Content Big Ideas 2. Writing Skills 3. Procedural Activities <p>*very few changes to Poetry curriculum</p> <p style="text-align: center;">For the following reasons:</p> <ul style="list-style-type: none"> • Teacher knows what is best for her class • Based on her knowledge of students' abilities and skills • Time constraints <p style="text-align: center;">In general:</p> <ul style="list-style-type: none"> • More likely to make modifications than omissions • More likely to make acceptable modification choices • Very likely to self-report implementation of intervention like the observer 	<p style="text-align: center;">STATISTICS</p> <p>Range of fidelity scores: 70-85</p> <p>Average Observed FOI score: 78</p> <p>Average Self-reported FOI scores: 86</p> <p style="text-align: center;">More likely to make modifications: (in order of percent occurrence) to the following Design Principles</p> <ol style="list-style-type: none"> 1. Respectful task based on student readiness, interest, or learning profile 2. Facilitation of tools and materials of the discipline 3. Flexible groups/Multiple modes of grouping 4. Depth: Exploration of details, patterns, and rules 5. Literary focus on reading comprehension activities <p style="text-align: center;">and a few Content Critical Component</p> <ol style="list-style-type: none"> 1. Procedural Activities 2. Content Big Ideas 3. Vocabulary <p>*very few changes to Poetry curriculum</p> <p style="text-align: center;">For the following reasons:</p> <ul style="list-style-type: none"> • Teacher knows what is best for her class • Unintentional omission/Forgot • Time constraints/ Teacher knowledge of student ability <p style="text-align: center;">In general:</p> <ul style="list-style-type: none"> • Twice as likely to make modifications than omissions • More likely to make unacceptable modification or omission choices • Wide range of Design Principal and Critical Component changes in each 	<p style="text-align: center;">STATISTICS</p> <p>Range of fidelity scores: Below 70</p> <p>Average Observed FOI score: 54</p> <p>Average Self-reported FOI scores: 73</p> <p style="text-align: center;">Most likely to omit or make unacceptable modifications: (in order of percent occurrences) to the following Design Principles</p> <ol style="list-style-type: none"> 1. Depth: Exploration of details, patterns, and rules 2. Flexible groups/Multiple modes of grouping 3. Continuous assessment measures 4. Big Ideas/Quality Curriculum 5. Respectful task based on student readiness, interest, or learning profile <p style="text-align: center;">and Content Critical Component</p> <ol style="list-style-type: none"> 1. Reading skills 2. Big Ideas/Procedural Activities 3. Communicating Findings/Presenting <p>*very few changes to Poetry curriculum</p> <p style="text-align: center;">For the following reasons:</p> <ul style="list-style-type: none"> • Teacher perceptions and expectations • Based on her knowledge of students' abilities and skills • Teacher knows what is best for her class <p style="text-align: center;">In general:</p> <ul style="list-style-type: none"> • More likely to make omissions than modifications • Twice as likely to make omissions than High or Moderate levels • Twice as likely to make unacceptable choices than High or Moderate levels