

2011

Exploring the Relationship Between Fidelity of Implementation and Academic Achievement in a Third-Grade Gifted Curriculum: A Mixed-Methods Study

Lisa Foster

Liberty University, lafoster@liberty.edu

Amy Azano

Tracy C. Missett

Carolyn M. Callahan

Sarah Oh

See next page for additional authors

Follow this and additional works at: http://digitalcommons.liberty.edu/educ_fac_pubs

Recommended Citation

Foster, Lisa; Azano, Amy; Missett, Tracy C.; Callahan, Carolyn M.; Oh, Sarah; Brunner, Marguerite; and Moon, Tonya R., "Exploring the Relationship Between Fidelity of Implementation and Academic Achievement in a Third-Grade Gifted Curriculum: A Mixed-Methods Study" (2011). *Faculty Publications and Presentations*. 232.

http://digitalcommons.liberty.edu/educ_fac_pubs/232

This Article is brought to you for free and open access by the School of Education at DigitalCommons@Liberty University. It has been accepted for inclusion in Faculty Publications and Presentations by an authorized administrator of DigitalCommons@Liberty University. For more information, please contact scholarlycommunication@liberty.edu.

Author(s)

Lisa Foster, Amy Azano, Tracy C. Missett, Carolyn M. Callahan, Sarah Oh, Marguerite Brunner, and Tonya R. Moon

Exploring the Relationship between Fidelity of Implementation and Academic Achievement in a Third-Grade Gifted Curriculum: A Mixed-Methods Study

Amy Azano, Tracy C. Missett, Carolyn M. Callahan, Sarah Oh, Marguerite Brunner, Lisa Foster, and Tonya R. Moon

Introduction

With educational practices currently receiving deep scrutiny, researchers increasingly focus on the impact of research-based curricula and teacher practices on student achievement. The need to attend to research-based practice extends to the field of gifted education where research should be directed toward understanding and measuring classroom environments in which gifted children learn, as well as the educational practices which best respond to the learning characteristics of these students (Plucker & Callahan, 2008).

In order to measure the impact of research-based curricula designed for the gifted classroom, researchers must consider the degree to which teachers who implement such curricula do so with fidelity to the research-based design (Lillehoj, Griffin, & Spoth, 2004; Lynch & O'Donnell, 2005). Logically, if teachers do not implement research-based curricular interventions as designed, measured outcomes cannot be attributed to the effectiveness of the intervention. However, few studies measure fidelity of implementation, or “the extent to which delivery of an intervention adheres to the protocol or program model originally developed” (Mowbray, Holter, Teague, & Bybee, 1999, p. 315). Even fewer explore the ways in which teachers’ beliefs, expectations, and perceptions influence their adherence to an intervention (Kennedy, 2004).

The purpose of this study was twofold. First, the ways in which teacher beliefs, expectations, and perceptions influenced teacher implementation of research-based curriculum were examined. Then, follow-up analyses of student achievement test scores were conducted to

assess the relationship between teachers' fidelity to the curricular model and student performance.

Research-based Curriculum and Context

The research described here is embedded within a Javits Grant funded study conducted by the National Research Center on the Gifted and Talented (NRC/GT) to develop empirical and descriptive understandings of effective curriculum for gifted students. This larger study is called "What Works in Gifted Education?" (WWIGE). As described more fully below, WWIGE studies the impact on student achievement of two challenging units in language arts/reading developed by experts in the field of gifted education from the University of Virginia for third grade self-contained and pull-out gifted classrooms.

The units are based on three well-known models in gifted education: (a) Tomlinson's Differentiated Instruction Model; (b) Renzulli and Reis's Schoolwide Enrichment Model (SEM); and, (c) Kaplan's Depth and Complexity Model. The Differentiated Instruction Model articulates and organizes curricular and instructional approaches designed to meet the needs of individual learners (Tomlinson, 2010). SEM encourages talent development and deliberately and systematically provides students with the skills, knowledge, and attitudes necessary for authentic work and original products (Renzulli & Reis, 2010). The Depth and Complexity model provides a vocabulary and strategies for designing rich and challenging curriculum specifically for gifted learners (Kaplan, 2005).

For purposes of the WWIGE study, distinct components of these models were synthesized and integrated into the CLEAR (**C**hallenge **L**eading to **E**ngagement, **A**chievement and **R**esults) Curriculum Model. The CLEAR Curriculum Model was developed to incorporate and reflect best practices in gifted education and to provide a framework for designing high-

quality, authentic curriculum appropriate for diverse learners, including students identified as gifted and those capable of advanced work.

The CLEAR Curriculum Model served as the theoretical and philosophical underpinning for two language arts/reading units. One, “The Magic of Everyday Things,” focuses on poetry. The other, “Exploration and Communication,” focuses on expository, nonfiction text, and research skills (referred to as the “poetry” and “research” units, respectively). Lessons for both units were designed for 45 – 60 minute instructional blocks. Both units are fully differentiated to challenge gifted learners with varying skills, interests, and readiness levels in a variety of settings. Additional principles of differentiation, such as formative and ongoing assessment and flexible grouping, are applied throughout the units.

In the broader study, the impact of the two units will be based on measured outcomes of students exposed to the intervention using standards-based assessments and structured performance assessments. In order to evaluate whether these measured outcomes are attributable to the intervention, both the broader study and this more limited study consider fidelity of implementation, specifically how beliefs and expectations influence a teacher’s adherence to the research-based curriculum.

Review of the Literature

Fidelity of Implementation

Educators are increasingly called upon to utilize a variety of curricular and instructional approaches to meet the diverse needs of students and improve student achievement. The *No Child Left Behind Act* (2001) requires educators to “use only research-based teaching methods and programs ‘proven’ to be effective” (O’Donnell, 2008, p. 35). Therefore, researchers and educators should be concerned about the extent to which the educational interventions they

design, implement, and adopt are evidence-based (Lynch & O'Donnell, 2005; Mowbray et al., 2003; Slavin, 2003).

In order to determine the effectiveness and impact of a research-based curricular intervention, assessing the degree to which its implementation adheres to the model of curriculum and instruction inherent in the research design becomes important (O'Donnell, 2004), a concept referred to as fidelity of implementation (FOI). While there are many competing definitions, we broadly address FOI as the extent to which delivery of an intervention adheres to the protocol or program model originally developed (Mowbray et al., 2003). Assessing FOI in the context of any intervention study is important because “[f]ailure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation” (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001, p. 39).

In studies assessing FOI, some empirical evidence suggests significant correlations between the extent to which intervention studies are implemented with fidelity and the level of treatment outcomes (Blakely et al., 1987; Dane & Schneider, 1998; Ruiz-Primo, 2006). While most of these studies have taken place in the context of public health and counseling research (O'Donnell, 2008), FOI studies evaluating the effectiveness of K-12 curricular interventions have been increasingly reported (Lillehoj et al., 2004; Lynch & O'Donnell, 2005; O'Donnell, 2008). When evaluating teacher fidelity in the context of such research-based curricula, a five-component framework has been proposed to include: adherence, exposure, program differentiation, quality of delivery, and participant responsiveness (see Table 1) (Lynch & O'Donnell, 2005; O'Donnell, 2008). Given the “compelling need to understand how curricula, instruction, and student diversity affect student achievement in the K-12 classroom” (Lynch & O'Donnell, p. 2), fidelity studies examining the effectiveness of K-12 core curriculum

interventions interpreting the impact of teacher FOI on outcome measures warrant attention and have been encouraged (O'Donnell, 2008).

Notably, in order to assess FOI, it is not necessary to evaluate all five components. Researchers should attend to those components that are of interest to their study (Lillehoj et al., 2004; Lynch & O'Donnell, 2005), and few report measuring all five. For example, Lynch and O'Donnell (2005) monitored adherence, quality of delivery, and participant responsiveness in evaluating teacher fidelity during the implementation of a research-based middle school science curriculum. Similarly, Lillehoj et al. (2004) focused on adherence and quality of delivery while assessing fidelity in the context of a middle school problem behaviors study.

For purposes of our study, the components of adherence and quality of delivery are of primary interest, specifically how teacher beliefs and experiences influence these two fidelity components. The research-based curriculum conforms to the gifted education field's current knowledge of practices enjoying the greatest support in the literature. As such, the curriculum is distinct from standard curriculum obviating the need to examine program differentiation. Moreover, the researchers were not able to fully assess and measure participant responsiveness or student engagement during delivery of the curriculum. Finally, investigation of adherence and quality of delivery precluded the need for inquiry into exposure as a distinct component. As indicated in Table 1, exposure (as one dimension of fidelity) looks at the number of lessons taught, time spent, and concepts taught. These variables were considered in our examination of adherence, as well as teachers' reasons for their decisions about adherence, and how these factors influenced delivery.

Best Practices in Gifted Education

To reflect best practices in the field, the National Association for Gifted Children (NAGC) issued the *Pre-K-Grade 12 Gifted Programming Standards* (the *Standards*) (NAGC, 2010). The stated purpose for the *Standards* is to help “teachers and other educators in Pre-K-12 settings to be effective in working with” gifted learners by helping them “understand the characteristics and needs of the population for whom they are planning curriculum, instruction, assessment, programs, and services” (NAGC, 2010, p. 1).

The *Standards* are based on literature identifying well-supported characteristics of gifted students (NAGC, 2010). These characteristics include, among others: a desire to focus on the essential knowledge, skills, and “big ideas” within a unit of study; a preference for open-ended and ill-defined academic experiences and tasks; the ability to proceed through curriculum at an individualized and accelerated pace; and, a preference for independent work and in-depth pursuit of topics of interest to the student (Renzulli & Reis, 2010; Tomlinson, 2005). While these characteristics are broadly associated with the gifted learner, educators must also recognize that “gifted learners themselves are anything but formulaic,” (Tomlinson, p. 160) and they present with varying strengths, interests, and needs that require differentiation.

Enrichment and acceleration are two principle approaches to differentiating gifted curricula and both are incorporated into the *Standards* (NAGC, 2010; Renzulli & Renzulli, 2010; VanTassel-Baska & Little, 2011). Enrichment broadly refers to broadening the scope of the curriculum beyond what is typically covered and/or increasing the depth of content studied (Reis & Renzulli, 2003). Enrichment in the context of rich curriculum allows students to analyze the layers of a discipline or content by going from concrete to abstract concepts, and from familiar to unfamiliar academic experiences (Reis & Renzulli, 2003; Tomlinson, 2005). It is distinguished from repeating material already taught and learned and from those classroom activities that fail

to provide students the “opportunity to delve into advanced issues and content” (Reis, 2003, p. 195).

Acceleration, on the other hand, broadly refers to faster academic progress that matches the “level, complexity, and pace of the curriculum to the readiness and motivation of the student” (Colangelo, Aussoline & Gross, 2005, p. *xi*). Although acceleration is often associated solely with more radical academic interventions such as grade-skipping and early entrance, it can and should be utilized in a variety of ways through curriculum compacting and flexible grouping strategies to meet the range of student skills and readiness levels within a classroom (NAGC, 2010; VanTassel-Baska & Little, 2011).

SEM is a prominent model for enrichment programming with a deep theoretical basis (Reis & Renzulli, 2003). Recognizing the diversity of skills and interests that gifted students possess, SEM focuses on the development of creative-productive giftedness through a variety of enrichment activities emphasizing “the use and application of information (content) and thinking skills in an integrated, inductive, and real-problem oriented manner” (Renzulli & Reis, 2010, p. 142).

Tomlinson’s Differentiated Instruction Model also enjoys wide use in educational programming for gifted students. Tomlinson (2010) summarizes this model as follows: “At the core of the classroom practice of differentiation is the modification of four curriculum-related elements – content, process, product, and affect – which are based on three categories of student need and variance – readiness, interest, and learning profile” (p. 15). Like SEM, the Differentiated Instruction Model incorporates a variety of curricular and instructional strategies to meet the diverse needs and profiles of gifted learners. They include focusing on the “Big Ideas” and concepts in a discipline, matching the pace, degree of challenge, and interests of each

student to instructional tasks, and allowing students to create individual products that reflect learning (Tomlinson).

Kaplan's Depth and Complexity Model provides for using standards-based curriculum as the foundation for creating additional challenge and extended understanding by integrating 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) (Kaplan, 2005).

Finally, the *Standards* also state that assessment is integral to the decision-making and teaching of gifted students (NAGC, 2010). Thus, the *Standards* urge educators of the gifted to use the results of formative, performance-based, and pre-assessments to adjust instruction and planning and to enhance ongoing learning progress (NAGC, 2010).

Teacher Expectancy Theory

Teacher expectancy theory generally refers to teachers' differential expectations about the learning capability of students within a classroom. These expectations are often based on perceptual bias rather than on students' actual performance or capabilities (Good & Nichols, 2001; Kuklinski & Weinstein, 2000). Within teacher expectancy theory is the notion that teachers often "hold a deficit-oriented framework when considering the characteristics of the primary-grade learner" (Moon & Brighton, 2008, p. 474). In the gifted education classroom, teachers holding a deficit-orientation attend to "remediation for these deficits before suggesting any enrichment, acceleration, or other gifted intervention strategies for their evident strengths" (Moon & Brighton, 2008, p. 474). The results of studies investigating the effect of teacher expectancy suggest that students often confirm the expectations teachers hold for them, and teacher beliefs about the abilities of their students appear to influence student achievement (Kolb

& Jussim, 1994). Alternatively, teacher expectations may produce perceptual biases when “teachers perceive, evaluate, or remember their students’ behaviors in ways that are consistent with their erroneous beliefs” (Kolb & Jussim, 1994, p. 28). Consequently, when teachers have low expectations for their students’ capabilities, student achievement is lower.

Although teacher expectancy theory is generally applied to differential expectations about the learning capability of minority or of high- versus low-achieving students and the resulting differential treatment of those students within the same classroom, some research posits this theory may have applicability to the expectations applied to whole classrooms (Reis, 2003). However, this research is scant. In addition, one would expect stereotypes of gifted learners to be reflected in high teacher expectancies for classrooms of gifted students. Pilot studies using the curriculum in our study suggested that universally high expectancies for gifted students are an unwarranted assumption. Thus, the impact of teacher expectancies in the gifted classroom, particularly when research-based curricular interventions are being implemented, merits consideration.

Using teacher expectancy theory as a conceptual framework, we sought to understand teacher beliefs about and expectations of student capabilities in the context of the gifted education classroom. We further sought to understand how those beliefs and expectations affected teacher adherence to and delivery of the educational practices and strategies reflected in a research-based advanced language arts curriculum. Finally, we sought to determine whether or not adherence and quality of delivery, as influenced by teacher expectations and beliefs, are related to measured student outcomes.

Method

The researchers employed a sequential mixed-methods research design, beginning with the qualitative component of the study. In doing so, researchers relied primarily on interpretivist methods and techniques (Erickson, 1986) for the major purpose of understanding the experiences and beliefs of the participants and how these beliefs and expectations influenced their adherence to and delivery of the research-based curriculum (Maxwell, 2005). Secondly, we used quantitative analysis to illuminate the qualitative component of the study and further understand whether the degree to which a teacher exhibited FOI (i.e., adherence and quality of delivery) is associated with student outcomes on researcher-designed, criterion-referenced post-assessments.

Participant Sample

Teachers were randomly assigned to experimental and treatment classrooms in the larger WWIGE study. Participants in the experimental condition included 55 teachers with a total of 740 students in urban, suburban, and rural schools across 10 states: Virginia, Maryland, South Carolina, Colorado, Kentucky, Florida, Pennsylvania, Connecticut, Georgia, and Texas. The units were implemented in self-contained and pull-out classrooms for identified gifted students during the 2009-2010 school year. (An additional 32 classrooms, which included 520 students, served as a comparison group for the broader study.) The qualitative analysis of data from the 55 experimental classrooms resulted in these teachers being assigned to one of three groups: high fidelity, medium fidelity, and low fidelity. The basis for this assignment is described below.

After qualitative analysis of observation and interview data from the 55 experimental teachers, researchers used another purposive sampling strategy, maximum variation sampling, to understand a wide range of teacher fidelity levels, and their impact on another phenomenon of interest, namely student outcomes (Creswell, 2007; Miles & Huberman, 1994). Maximum variation sampling is appropriate in emergent or sequential methodological approaches such as

the one employed here where conclusions drawn at one phase of a study (the influence of teacher beliefs and expectations on fidelity) can then be used to inform further direction of the study (student outcomes) (Creswell, 2007). This sampling strategy resulted in a reduced sample size of 26 teachers, 12 of whom were categorized as high fidelity and 14 as low fidelity, for subsequent quantitative analysis. This methodological approach allowed the researchers to more precisely explore the relationship between teacher fidelity and student outcomes in a manner consistent with the purposes of maximum variation sampling.

Data Sources

Qualitative sources of data included observations in a field setting while teachers implemented the reading units and intensive, follow-up participant interviews. Researchers observed and interviewed 96% of teachers in the experimental condition at least once; 67% were observed and interviewed on three separate occasions. Observations were recorded on a semi-structured protocol specifically designed to note where teachers adhered to, or did not adhere to, the lessons. Following the observations, researchers interviewed participants using semi-structured interview protocols to gain an understanding of the fidelity and instructional choices teachers made during implementation. These interviews were recorded and transcribed for later analysis.

Quantitative data sources included pre- and post-assessments. To ensure equity of achievement levels prior to intervention, student performance on the Iowa Tests of Basic Skills (ITBS) Reading Comprehension subtest (Level 9, Form A) was assessed at the beginning of the 2009-2010 school year. The Reading Comprehension subtest measures how students derive meaning from a variety of reading passages representing narrative, poetry, and nonfiction material from science and social studies. Reliability estimates of the Reading Comprehension

subtest are generally in the range of .95 (Hoover et al., 2003). For the student outcome measurement, two criterion-referenced post-tests with 35 items each were constructed by the research team and piloted at treatment sites in different states in order to determine appropriate level of difficulty and discrimination for each item. Post-tests were administered in May 2010 at both treatment and comparison sites. Cronbach's α and Spearman-Brown reliability estimates ranged from .72 to .79 for both of the tests.

Data Analysis

Qualitative Data Analysis. A research team, including a research scientist and two graduate research assistants, analyzed the qualitative data. The team members are experienced educators and have a specialty area in gifted education, qualitative methodology, and reading instruction. Analytic induction (Erickson, 1986), which included reading the data corpus (observations and interviews), coupled with initial memo writing on themes emerging from the data occurred in three stages.

During the first stage of the data analysis, the researchers read the data corpus to “develop tentative ideas about categories and relationships” (Maxwell, 2005, p. 96). All initial categories synthesized the observational data with its companion interview data. This allowed the research team to compare the researcher-generated observational data to the related teacher interview data. For example, if a researcher observed that a teacher had omitted the formative assessments incorporated into the intervention, or had failed to employ the grouping strategies set forth therein, the research team categorized those practices as lacking adherence. The research team then reviewed interview data to ensure that these non-adherent practices or modifications to the intervention were discussed and explained during the follow-up interview. This process allowed the researchers to simultaneously identify those components of the

intervention from which the participants deviated while understanding the teacher beliefs undergirding the deviations.

The initial reading of the data allowed the researchers to capture both adherence to the intervention as well as a set of beliefs and experiences that indicated a teacher's adherence and quality of delivery. These beliefs and experiences, which provided initial coding categories, included: a teacher's perception of the amount of time she¹ had to implement the intervention (time); the expectations she held for students' abilities to understand and learn the content and skills in the intervention (teacher expectations); the degree of autonomy and support she felt while implementing the intervention (autonomy); and, the degree to which she felt professionally competent in the areas of poetry and research instruction (professional expertise).

To promote coder agreement, research team members independently coded the interview transcripts of three teachers. They then met to compare the application of codes and to refine their understanding about code applicability. The team divided the remaining data for coding with frequent interaction and feedback across the team. This exercise was repeated until coders aligned on the application of the codes. Periodically the data were re-coded by a different team member to check for consistent coding. In all, three iterations of the coding process were required to reach inter-reliability of .90.

The research team found that each coding category (time, autonomy, teacher expectations, and professional expertise) represented a continuum reflecting a range of beliefs and expectations that could facilitate or serve as a barrier to adherence and quality of delivery. For example, the researchers observed that teachers devoted varying amounts of time to implementing lessons contained in the unit. During the interviews, researchers learned that the amount of time a teacher spent on a lesson depended on her perception of time available. The

¹ All participants in the study were women.

participants' perception of time as captured in interview data varied from "I have as much time as I need to teach these units just as you asked" (teacher interview, March 23, 2010), to "I really have to be creative in finding the time I need to teach these lessons because I have so many students" (teacher interview, October 14, 2009), to "of course my whole year had to be adjusted because we had a whole spate of tests that we had to give, that kind of thing and that's made it more difficult to implement the units" (teacher interview, April 16, 2010), to "I just don't have time to do it all" (teacher interview, October 27, 2009). The teachers' varying perspectives of time which differentially impacted their adherence were ultimately coded under a perception of time category.

Similarly, researchers observed that some participants adhered strictly to the lessons with virtually no modifications. Participants' beliefs about autonomy captured from interview data ranged from "since I agreed to participate in this study and my principal supports me, I taught the units exactly as you asked" (teacher interview, March 16, 2010), to "although I wouldn't necessarily have chosen to do this, my district coordinator has given me all I need to make it happen" (teacher interview, October 9, 2009), to "if my principal hadn't told me I have to participate in this study, I would not have taken it on" (teacher interview, March 26, 2010). The teachers' varying beliefs about their autonomy were ultimately coded under an autonomy category.

Another example is seen in the degree to which teachers utilized grouping and differentiation strategies incorporated into the intervention. The researchers observed that teachers adjusted the pace of instruction to the readiness levels of individual students in varying degrees with some adopting a dominant "one size fits all" pacing strategy; others used strategies allowing students to move through the curriculum at their own pace. When asked about their

spacing choices, teachers offered varying explanations ranging from “I noticed some students were having problems” so all students repeated an activity (teacher interview, February 16, 2010), to “since almost all of my students understood the concepts, I let them move forward while I conferenced with two girls who I had to get up to speed” (teacher interview, December 4, 2009). Generally, these comments reflected differential expectations held by teachers for their students’ ability. The teachers’ varying expectations were ultimately coded under a teacher expectations category.

After developing the initial set of categorizing codes reflecting teacher beliefs and experiences which contributed to adherence and quality of delivery, the research team completed a second reading of the qualitative data corpus to further develop codes. Again, the team reviewed both observational and interview data to confirm that observations regarding fidelity were corroborated or explained by teacher interview statements. The researchers looked for confirming and disconfirming evidence of the codes to ensure that the analytic process plausibly captured the beliefs and experiences that contributed to FOI. The team then subjected the data to further categorizing strategies and thematic analysis.

In the second stage, the research team repeated the steps described above. They then inductively developed, through a close “open coding” of the data to describe more substantively the theoretical categories, a more refined set of codes reflecting more differentiated beliefs and experiences contributing to adherence and quality of delivery (Maxwell, p. 97). These sub-codes helped further delineate the initial coding categories and became the final codes reflected in Table 2.

Researchers applied the codes to observational and interview data. For example, in coding the observational data, we observed that Mrs. Robbins² adhered to the lessons in the units consistently. When Mrs. Robbins was asked to explain the reasons for not modifying the lessons during follow-up interviews, she replied: “I’ve been trying to stick to the lessons exactly as they’re written because it’s research, and I signed on to this” (Robbins interview, March 16, 2010). This data received the “AUT – P” code because it reflected Mrs. Robbins’ beliefs about her autonomous role as participant.

As another example, in coding the observational data, we found that Mrs. Baker made extensive modifications to the intervention in, among other things, the grouping strategies she employed. Specifically, when the intervention prescribed small group or pair share activities, Mrs. Baker consistently utilized whole group practices. In follow-up interviews seeking to understand these modifications, Mrs. Baker explained “my kids really want to hear from each other. They just don’t have fun if they don’t get to share with each other. And if they’re not having fun, they’re not going to learn” (Baker interview, March 8, 2010). Because these statements indicated Mrs. Baker’s expectations that her students were less likely to succeed without modifications, this data received TE – F and TE – S codes.

Finally, in a third stage of qualitative analysis, we found that teachers could be grouped into three categories: high adherence, mixed adherence, and low adherence. From the exhaustive coding and inductive analytic process described above, teachers were grouped according to their overall fidelity to the lessons as written and to the theoretical strategies embedded within. Researchers relied on observational data that found broad support in interview statements in making these decisions. See the findings section for details. This finding is explicated here to provide the context of the analysis of student outcome data which follows.

² Pseudonyms have been assigned for all people and places.

Multilevel Analysis of Student Outcome Data. Once the qualitative data analysis was complete, the researchers used the high and low teacher categories to further inform the quantitative data analysis to examine the association between teachers' adherence to the curriculum and student outcomes.

Using maximum variation sampling (Creswell, 2007; Miles & Huberman, 1994) as a complimentary method to illuminate the broadest differences in overall beliefs and experiences between teachers in the high adherence group and teachers in the low adherence group, student outcomes from the high adherence teachers' classrooms were compared to outcome measures from low adherence teachers' classrooms. While the qualitative analysis resulted in the placement of 12 teachers in the "high adherence" group and 14 in the "low adherence" group, ultimately, the sample for the quantitative component of the study included 11 teachers from the "high adherence" group, and 13 from the "low adherence" group due to the fact that one teacher from each group did not return post-test data for quantitative analysis. These two groups were compared for purposes of analyzing student outcome data. Consistent with the principles of maximum variation sampling, "mixed adherence" teachers were not included in the quantitative analysis.

Multilevel analyses for both units allowed the nested nature of the data set to be taken into account and prevented issues with aggregation bias and the misinterpretation of standard errors in the analysis (Raudenbush & Bryk, 2002). All the analyses used *Mplus* 6 as it allows maximum likelihood parameter estimates with standard errors and chi-square test statistics that are robust to non-normality and non-independence of observations in clustered data (Harnqvist et al., 1994). Prior to data analysis, intraclass correlation (ICC) and design effect (DE) statistics were calculated to decide whether multilevel modeling was necessary. A significant variance

across classrooms was observed with ICC ranging from 0.15 to 0.33 for both units. In addition, a higher rate of the DE ranging from 3.02 to 5.87 also supported the need for multilevel modeling for outcome measurement for both units. In order to examine the relationship between teachers' adherence to the curricular intervention and student 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 curriculum. 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 scores were entered after grand-mean centering as a proxy for previous unaccounted influences and other categorical variables were left uncentered.

The final models for analysis appear below:

Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Grand_ITBS})_{ij} + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Teacher Adherence})_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Findings

Qualitative findings

Data analysis provided insights into the ways the beliefs and experiences of teachers influence fidelity during the implementation of the research-based curriculum. In other words, these beliefs and experiences influence both the degree to which teachers adhere to the units as written, as well as the degree to which the quality of their delivery comports with the theoretical and pedagogical ideals contained in them.

Specifically, the research findings revealed that a teacher's sense of autonomy and professional expertise, expectations of student capability, and perceptions of time all interact with each other to either facilitate or constrain adherence and quality of delivery. These interactions are dynamic in that each tends to exert an influence on the others. However, the interactions among beliefs and experiences varied across teachers so that no one indicator influenced adherence and quality of delivery in exactly the same way for all teachers. Nevertheless, indicators of fidelity including teacher perceptions of autonomy, perceptions of time, expectations of student capability, and beliefs about their professional experience support the following six general assertions:

1. Adherence falls along a continuum of high adherence to low adherence.
2. Quality of delivery falls along a continuum of high quality to low quality.
3. Teacher beliefs about their degree of professional autonomy fall along a continuum of high autonomy to low autonomy. The more professional autonomy a teacher experiences, the higher the teacher's adherence and quality of delivery tend to be.
4. Teachers' perceptions of the instructional time they have to devote to the intervention fall along a continuum of high perceptions to low perceptions. The higher a teacher's perception of available instructional time, the higher the teacher's adherence and quality of delivery tend to be.
5. Teacher beliefs and expectations fall along a continuum from high to low. The higher a teacher's expectation for student performance or the higher the teacher's belief about her students' abilities, the higher the teacher's adherence and quality of delivery tend to be.

6. Teachers' beliefs about their own professional expertise fall along a continuum of high beliefs to low beliefs. The higher a teacher's belief about her own professional expertise, the higher the teacher's adherence and quality of delivery tend to be.

High Adherence Teacher. Overall, a combination of teacher beliefs reflecting high levels of autonomy, high expectations of student performance and belief in students' capability, perceptions of sufficient time to implement the units, and/or strong beliefs about their own professional expertise interacted with each other to indicate high adherence and high quality of delivery. We refer to these participants as high adherence teachers as their beliefs and experiences combined to facilitate fidelity. Twelve participants can be described as adhering closely to the research-based curriculum and delivering the WWIGE units with high quality. High adherence was indicated by overall fidelity to the units as written as observed by the researchers and as understood by teachers in interviews. High quality was indicated by the use of best practices including (1) appropriate accelerative strategies that set the pace of instruction to the readiness level of students, (2) appropriate enrichment strategies that were directed toward individual student interests, and (3) integrated instructional strategies that differentiated based on individual student need. Incorporating specific interview data, the following vignette illustrates the typical experiences and beliefs of a high adherence teacher.

I've been a gifted resource teacher for six years. Our students are identified for gifted program services in second grade based on their abilities test scores. This year I have 12 third grade students. I see them for an hour and a half every day. I usually do language arts three days per week and math and critical thinking activities two days per week. My students get most of their core third grade curriculum instruction in their regular education class.

These reading units fit perfectly in my program. They are quite different from the language arts instruction my students get in their regular classroom where the kids spend a lot of time preparing for standardized tests, which I don't have to worry about. For my students, this is hardly challenging or engaging. So, when my students come to me I try to provide academic experiences that challenge them, and that allow them to go more deeply into areas of interest to

them. They are very curious and thoughtful students who I can push. There's really almost nothing I can't give them.

I feel like one of the luckiest teachers in the world. My principal and district coordinator really trust me to run my program as I think is best. I have everything I need and more from our librarian, our technology lab assistant, and the classroom teachers. They respect the time I have with my students and always make themselves available when I need them.

When I heard about the research you were doing, I thought it would be a great fit for my students. After I read the units, I was so impressed with the high level and strong quality that I immediately wanted to be a participant in the study. I've taught poetry and research skills before so I was comfortable teaching this material. I jumped right on board and proposed participating in the study to my district coordinator. I described both units, and I explained how much skills development and critical thinking were included in them. Without hesitation, my coordinator and principal said go for it. Ever since I committed to the study, they've completely supported me. My students love the units also.

I can usually fit at least two lessons from your study into our week. While there are about 24 students in their regular classrooms, I only have twelve. It makes instruction much more manageable. Although my students are all identified gifted, they still have varied levels of skills and abilities, so I do need to differentiate for them, even in these units. I've found the formative assessments in the units to be really helpful, and have discovered that the groups are fluid.

My expectations for my students' capabilities are high. For example, when students raise their hands and say, "I need help" or "this is so hard," I don't help them right away. Instead I'll say "you've been at it for one minute. It's too soon to give up" or "it's supposed to be hard, but I know you can do it." I don't hesitate to let my students know when their work product is strong. I say things like "What a great metaphor. I've never thought of a snake in that way but I sure will now." I also let them know when they haven't reached their potential. I'll tell them, "I know you can describe that scene with more detail. Use your senses. Keep going." Of course, sometimes I have to give one or two students a little more attention because they're all at different levels. But in those instances I schedule some one on one time while the rest of my students keep moving ahead.

I taught the Explorers unit first. My students really had to stretch with some of the independent work, but they got through it. Gifted kids shouldn't expect school to be without challenges. I can't remember making any modifications to the unit, except once or twice asking the kids to finish some of the research at home if they didn't finish during class time. I'll be honest here. I really considered modifying the poetry unit because when I read the poems, I thought they were so abstract and far off from their experiences that they'd get very little out of them. I even had supplemental materials ready to go. However, since I signed on to your study, I felt committed to sticking to the plans and I didn't use that stuff.

I'm glad I didn't change the lessons and I must say I've been really impressed. With everything I've taught them they've been like sponges. They're using the vocabulary and strategies appropriately and developing reading skills. I realized from teaching the poetry unit that when you give learning back to the kids, they'll

understand much more than we might otherwise expect. Every day I come to work and I get to be amazed at the work my students produce. My students are in the gifted program for a reason. They're incredibly capable.

It is important to note that high adherence teachers exhibited varying beliefs about their levels of autonomy, expectations of student capability, perceptions of time, and professional expertise. Thus, for example, it was still possible for a teacher to express low beliefs about her professional expertise ("I have never taught poetry before and I was uncomfortable during that unit," teacher interview, March 14, 2010) and still be considered a high adherence teacher based on her overwhelmingly high expectations for student capability that more forcefully influenced her implementation of the intervention. In other words, high adherence teachers were those teachers for whom implementation of the intervention was more forcefully guided by high levels of adherence and quality of delivery indicators.

Mixed Adherence Teacher. Twenty-nine participants, or approximately half, can be described as exhibiting mixed adherence to the research-based curriculum and delivering the intervention with mixed quality. Overall, these teachers expressed varying beliefs about their respective levels of autonomy, expectations of student capability, perceptions of time, and professional expertise. In turn, their varying beliefs interacted with each other to indicate mixed levels of adherence and quality of delivery to the research-based curriculum. We refer to these participants as mixed adherence teachers. The following vignette illustrates the typical experiences and beliefs of a mixed adherence teacher.

I teach 22 gifted third grade students. At my school, the gifted program is a self-contained model. To be identified, students have to have IQ and achievement scores at the 95th percentile. For low SES and minority students, we include students at the 90th percentile. Students also produce a portfolio which is taken into consideration when we identify for the program.

I first found out about the study from our district coordinator. She presented it to the three GRTs in the county and asked if we were interested. My co-workers were very interested, but they have fewer students. I said I'd do it

because the other two wanted to participate, but to be honest I probably would not have done so if it were left to me. I think the units are great and the kids have gotten a lot out of them. However, on top of everything else I have to do I'm really stretched. In fact, next year I'll probably have to go back to the classroom because my workload is so intense. My principal is great. However, his biggest concern is that our students pass the state tests. It's not what I call gifted education, and it takes away from what I think is appropriate curriculum for these kids, but I know we have to improve our test scores.

Even though I have my students all day, I don't have a lot of time to do these units on top of my other responsibilities. We have so much district testing we're still responsible for. I feel like I spend most of my day covering our state mandated curriculum and the testing that goes along with it before I get the chance to implement the research units. So when I'm running short on time, the first thing that I drop is these units. Sometimes a couple of weeks go by between lessons, which makes it necessary to cover some of the material again.

One way I try to get through the units is to blend the lessons with some of the curricula I'm already teaching. That way, I can kill two birds with one stone. For example, in our district we have an art history unit. I think it's useful to blend the art unit with the poetry unit because so many of the same concepts, like abstraction, imagery, and point of view, are in both. I can really enrich both units by integrating them. When we went to the art museum a few weeks ago, I was so happy to hear my students discussing what was abstract and what was concrete in the paintings. It showed me how transferrable your research units are.

We ran into a few difficulties getting library time, which impacted how I taught the Explorers unit. Our librarian is so swamped with her responsibilities that asking her to assist me on the research unit wasn't easy. We also have a limited number of computer terminals. Consequently, my students had to take turns doing their research, which slowed us down. Fortunately, our regular research curriculum had many of the strategies and skills the students learned in the Explorers unit. So, I skipped over some lessons in Explorers so we could get through it more efficiently. Plus, I'm very comfortable with the research process because I've taught the skills and strategies contained in the Explorers unit.

When I first started teaching the poetry unit, I really thought the material might be a bit too much for most of my students. The poems are so difficult that I worried they might be lost. Sometimes third grade students, even gifted students, need more accessible material. At first, I went ahead and added some poetry that was much simpler so I'd be sure they understood the concepts. However, as I went through the unit I was pleasantly surprised because these children – even when I thought this was going to be too hard for them – came up with some of the most insightful discussions. I was very pleased that even though the children might not get every meaning of every single poem they certainly did get a lot out of them and could pick out the imagery and poetic devices. So I bit my tongue and stepped back because they proved me wrong. Now, I stick to your poems.

As reflected in the above vignette, mixed adherence teachers by definition articulated a full range of beliefs indicating adherence and quality of delivery. For example, a mixed adherence teacher might have a combination of beliefs such as low autonomy and perceptions of time, high sense of professional expertise, and mixed expectations of student capability, all of which interact to yield a teacher showing mixed levels of adherence and quality of delivery. Thus, a mixed adherence teacher may have adhered highly to the grouping strategies in the units but not those addressing assessments.

Low Adherence Teacher. Fourteen participants can be described as exhibiting low adherence to the research-based curriculum and low quality of delivery. Overall, these teachers expressed broadly compromised beliefs on their respective levels of autonomy and expectations of student capability, perceptions of insufficient time to implement the units, and doubts as to their own professional expertise. In turn, their low beliefs interacted in such a way as to indicate low adherence and quality of delivery to the research-based curriculum. In other words, beliefs and expectations served as barriers to fidelity. Low adherence was indicated by the inconsistent implementation of numerous unit components as observed by the researchers and as acknowledged by these teachers in interviews. Low quality of delivery was indicated by the weak use of best practices reflected by (1) pacing instruction to the readiness level of the slowest learner(s), (2) limiting enrichment opportunities directed toward individual interests, and (3) inconsistent integration of differentiation strategies. The following vignette illustrates the typical experiences and beliefs of a low adherence teacher.

I'm a gifted resource teacher at my school. At my school, the gifted program utilizes both pull-out and push-in components. I work with students in third, fourth, and fifth grade who have been identified as gifted. To be identified, students have to have IQ scores at the 95th percentile, or at the 90th percentile if they have a strong teacher recommendation.

I work with third grade students during my pull-out time with them. I have 25 students. In theory, I see them twice a week for an hour. However, whenever there is an assembly or field trip, I lose my time slot with them. It seems like this class is not always the priority.

I'm not sure how my school became involved in this study. My district coordinator thought it was a great idea and told me I was going to do it. Although I really like the units, I feel like I don't really have the time to do them the way I've been asked. So, I probably would have preferred to not participate in your study. I'm just too buried. As you can see, I haven't even had time to fill in my teacher fidelity log.

I really love working with these kids. They are all very sweet and bright. They're still at the age where they're not afraid to be silly. They can work on material that is more difficult than most third grade students can handle. You would be amazed at how creative my students can be, and how they like to think deeply and seriously.

However, they're still children who want to enjoy what they're doing. In my opinion, some of the lessons are so challenging that I get concerned that they won't enjoy themselves. If my students aren't having fun with their work, they're just not going to grasp complex material. So I try to incorporate games and catchy activities into my lessons to keep them interested. For example, in the poetry unit I let my students bring in their favorite song lyrics to see if we could detect poetic devices in them. Plus, it gives each student time to share their interests with the class. They always want to hear what their classmates think. In fact, even though all the poetry lessons asked students to share their thoughts about the poems with only one partner, I preferred to do a whole group share. In addition to making class more fun, I believe sharing with the whole class helps reinforce the works we've done. Otherwise, they'll just tune out. I'll tell you, though, it really takes so much time to do it right. These units are taking a lot longer than I expected.

Now that we've gotten through the poetry unit, I must say the poetry in this unit is wonderful. I wouldn't necessarily have chosen this poetry for them because I thought it would have been too difficult. That could be a reflection of my own discomfort with poetry, I've just never taught poetry before. I didn't understand some of the poems you used, and I knew my students wouldn't either. I found I needed to take time to discuss the meaning, context, and vocabulary with them, and also did some author study work so they could learn more about the poets. I felt they needed more than just focusing on images the poem presented. But I was so surprised to see that they are capable of this. I still think they just need a little more explanation or a lower level of abstraction with practical examples of poetry most of the time before we move on to new lessons. So I supplement the poems in the lessons with my own that are more appropriate for third graders. If I think a worksheet is too hard for them, I'll help the group along with whole class instruction, or make one up on my own that I think works better for everyone. I also really tried to help kids come up with their research questions. I thought they would need guidance on that because it just seemed too complex for these kids.

I didn't use your formative assessments very much. Sometimes, the groups the assessments would probably create based on ability would just be too difficult in my class. Some of my students don't work well together and I'd rather have some of the higher students help some of the lower students. With so many kids in the class, I need to make the classroom work smoothly.

Again, it was possible for low adherence teachers to demonstrate adherence in some areas. However, their low beliefs in other categories more strongly influenced their overall adherence and quality of delivery. Thus, for example, a teacher with low perceptions of time and low autonomy could demonstrate low adherence and quality of delivery even if she believed she had high levels of professional expertise.

Quantitative findings

The multilevel analyses indicated that teachers' adherence to the curricular intervention has a significant association with student outcome measures after controlling for pre-test scores (on the ITBS Reading Comprehension subtest) for both units ($p < .05$). Students in high adherence teachers' classrooms scored 1.98 points higher in the poetry unit and 3.38 points higher in the research unit than students in low adherence teachers' classrooms. Model fit tests corroborate the significant association between teachers' adherence to the curriculum and student outcome for both units: $\Delta\chi^2_{df=1}=10.00$, $p < .01$ for the poetry unit and $\Delta\chi^2_{df=1}=6.12$ $p < .01$ for the research unit. As effect size indices for the teacher adherence to the curricular intervention, global effect size for each unit was calculated. Results indicated that a significant amount of the variance in student outcome (8.5% for the poetry unit and 12.3% for the research unit) can be explained by teachers' adherence to the curricular intervention. See Tables 3 and 4.

Discussion

The results of this study have important implications for the development and implementation of research-based curricula. Specifically, our results suggest that the beliefs and

experiences teachers bring to the classroom impact their instructional practices, notwithstanding the availability of quality curriculum. Here, the participants all had reading curricula available to them that conformed to best practices in gifted education. All were participants in a study measuring the impact of research-based curricula on student achievement. Nevertheless, the beliefs teachers held about their students, their own expertise, their autonomy, and the time required for implementing the units ultimately produced varying degrees of fidelity.

For example, the use of accelerative and enrichment strategies appropriate to gifted learners depended largely on the expectations teachers held for students. Thus, if a teacher expected that any of her students would struggle with the content and skills embedded within the units, she would often slow the pace of instruction for all students, or limit the individual choices available to them. In turn, these beliefs appear to have ultimately affected measured student outcomes. Conversely, if a teacher expected her students to be capable of engaging successfully with a challenging curriculum, she based her instruction on the readiness levels of individual students in a manner consistent with best practices. These results are consistent with prior studies investigating teacher expectancy effects on student outcomes (Moon & Brighton, 2008).

Often times barriers derived from external factors, such as instructional time with students or demanding district testing requirements. In other cases, low adherence derived from internal factors, such as a teacher's deficit-oriented expectations of her students or her own professional expertise. By comparing these two groups—"high adherence" teachers who emphasized facilitators for implementing the curriculum versus "low adherence" who emphasized barriers to implementation—we were able to determine that a teacher's level of adherence is related to student outcomes. While we do not suggest that these findings are causal in nature, it is an important point of further inquiry.

Implications and Directions for Future Research

The results relating to influences on teacher implementation of curricular interventions and the effects on outcomes are important for researchers who develop research-based classroom interventions as they support the notion that if teachers are not personally committed to participating in a study, lack the time and resources necessary to meaningfully implement an intervention with fidelity, or do not expect students to be capable of high quality curriculum, they will be less likely to adhere to research protocols. Because the results strongly suggest that teacher beliefs and expectations influence fidelity, creating and delivering professional development specifically addressing beliefs and expectations for research participants could be an important first step in mitigating these potentially adverse influences.

Data related to teacher characteristics, such as teaching experience, educational experience, years teaching gifted students, and coursework directed towards gifted learners, were collected. While one might assume that more classroom experience would yield enhanced adherence, we did not find that to be evident in this study (see Table 5). In fact, teachers in the “low fidelity” group had more total years experience teaching, teaching third grade, and teaching gifted learners. Similarly, teachers with advanced degrees did not represent a larger number in the “high fidelity” group (see Table 6). While these findings provide limited, descriptive insights, they do suggest that the variables influencing a teacher’s adherence and quality of delivery in a given curriculum are complex and multifaceted. While these characteristics were not the foci of this study, they do provide the opportunity for subsequent analyses to investigate the inter-relationships and the degree to which these personal characteristics influence teacher beliefs, expectations, fidelity of implementation, and student outcomes.

Finally, barriers to implementing research interventions with fidelity pose pragmatic challenges to researchers who endeavor to understand curriculum and instruction appropriate for diverse learners. The findings from this study will contribute not only to the field of gifted education, but also the research community's understanding of fidelity of implementation. Finally, by addressing how barriers to implementation may in turn influence student academic achievement, this mixed methods study will help to answer a breadth of questions that lead to empirical and descriptive understandings of "what works in gifted education" in the area of reading instruction.

References

- Blakely, C., Mayer, J. P., Gottschalk, R. G., Schmitt, No., & Davidson, W. S. (1987). The fidelity – Adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253-269.
- Colangelo, N., Assouline, S., & Gross, M. U. M. (2004). *A nation deceived: How schools hold back America's brightest students* (Vols. 1-2). Iowa City: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development, University of Iowa.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Dane, A. V. & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.
- Dumas, J., Lynch, A., Laughlin, J., Smith, E., & Prinz, R. (2001). Promoting intervention fidelity: Conceptual issues, methods and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine, 20*(1S), 38-47.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd Ed., pp. 119-161). New York: MacMillan.
- Good, T. L., & Nichols, S. L. (2001). Expectancy effects in the classroom: A special focus on improving the reading performance of minority students in first-grade classrooms. *Educational Psychologist, 36*, 113-126.
- Harnqvist, K., Gustafsson, J. E., Muthén, B., & Nelson, G. (1994). Hierarchical models of ability at class and individual levels. *Intelligence, 18*, 165-187.

Hoover, H. D., Dunbars, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., . . .

Shannon, G. P. (2003). *The Iowa Tests: Guide to research and development*. Itasca, IL: Riverside.

Kaplan, S. (2005). Layering differential curricula for gifted and talented. In F. Karnes & S. Bean (Eds.), *Methods and materials for teaching gifted students* (2nd ed., pp. 107-132). Waco, TX: Prufrock Press.

Kennedy, M. M. (2004). Reform ideals and teachers' practical intentions. *Education Policy Analysis Archives*, 12(13). Retrieved from <https://www.msu.edu/~mkennedy/publications/docs/Teaching%20Practice/PracticalIntentions/Kennedy%20Ideals%20and%20Intentions.pdf>

Kolb, K. J., & Jussim, L. (1994). Teacher expectations and underachieving gifted children. *Roeper Review*, 17, 26-30.

Kuklinski, M. R., & Weinstein, R. S. (2000). Classroom and grade level differences in the stability of teacher expectations and perceived differential teacher treatment. *Learning Environments Research*, 3, 1-34.

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior*, 31, 242-257.

Lynch, S., & O'Donnell, C. (2005, April). "Fidelity of implementation" in implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Moon, T. R., & Brighton, C. M. (2008). Primary teachers' conceptions of giftedness. *Journal for the Education of the Gifted, 31*, 447-480.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315-340.
- National Association for Gifted Children. (2010). *Pre-K-Grade 12 Standards*. Retrieved from <http://www.nagc.org/index.aspx?id=546>
- No Child Left Behind Act, Title IX, General Provisions, Part A, Section 9101, 20 U.S.C. § 7801 et seq. (2001). Retrieved from <http://www.ed.gov/policy/elsec/leg/esea02/pg107.html>.
- O'Donnell, C. L. (2004). Fidelity of implementation: Background, definitions, and components for measuring. Internal document: The George Washington University.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33-84.
- Plucker, J., & Callahan, C. (Eds.) (2009). *Critical issues and practices in gifted education: What the research says*. Waco, TX: Prufrock Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). London: Sage.

- Reis, S. M. (2003). Reconsidering regular curriculum for high-achieving students, gifted underachievers, and the relationship between gifted and regular education. In J.A. Borland (Ed.), *Rethinking Gifted Education* (pp. 186-200). New York, NY: Teachers College Press.
- Reis, S. M., & Renzulli, J. S. (2003). Research related to the Schoolwide Enrichment Triad Model. *Gifted Education International, 18*, 15-39.
- Renzulli, J. S., & Reis, S. R. (2010). The Schoolwide Enrichment Model: A focus on student strengths and interests. *Gifted Education International, 26*, 140-157.
- Ruiz-Primo, M. A. (2005, April). A multi-method and multi-source approach for studying fidelity of implementation. In S. Lynch and C. L. O'Donnell (Chair), "Fidelity of implementation" in *implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations*. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Slavin, R. (2003). A reader's guide to scientifically based research: Learning how to assess the validity of education research is vital for creating effective, sustained reform. *Educational Researcher, 60*(5), 12-16.
- Tomlinson, C. A. (2005). Quality curriculum and instruction for highly able students. *Theory Into Practice, 44*, 160-166.
- Tomlinson, C. A. (2010). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- VanTassel-Baska, J., & Wood, S. (2010). The integrated curriculum model (ICM). *Learning and Individual Differences, 20*, 345-357.

VanTassel-Baska, J., & Little, C. A. (2011). *Content based curriculum for high ability learners* (2nd ed.). Waco, TX: Prufrock Press.

Table 1

FOI Five-component Framework

Component	Definition
Adherence	Whether the unit is delivered as designed.
Exposure	Whether the number of lessons taught, the length of time spent by the teacher, and the type of concepts and skills delivered match the intent of the unit developer.
Program differentiation	The existence or absence of “critical features” that distinguish the unit from standard curriculum.
Quality of delivery	Whether a teacher implements a unit in a manner consistent with the “theoretical or pedagogical ideals” and techniques embedded within the unit.
Participant responsiveness	Whether the extent of student involvement or engagement in the unit’s activities matches the intent of the unit developer.

Table 2

Codes and Sub-codes for Qualitative Data Analysis

Codes	Sub-codes	Sub-codes Description
Autonomy (AUT) –		
Autonomy for this study broadly refers to the level of administrative support for and personal commitment to participate throughout implementation of the WWIGE study	Participant Commitment (AUT - P)	Degree to which teacher feels personally committed as participant in research
	Support (AUT - S)	Degree to which teacher believes she has what is needed in terms of materials, space, technology, etc.
	Administration/Principal (AUT - A/P)	Degree to which teacher believes she is supported by administration/principal throughout study
Perceptions of Time (logistics) (T) –		
Perceptions of Time refers to the degree to which a teacher felt she was or was not impacted by time constraints throughout implementation of the WWIGE study	Program Delivery (T - PD)	Degree to which pull-out v. self-contained program delivery impacted perceptions of time
	Testing (T - T)	Degree to which district/state testing requirements impacted perceptions of time
	Number of Students (T - NS)	Degree to which number of students impacted perceptions of time
	Library (T - L)	Degree to which library resources (librarian, books, technology) impacted perceptions of time

Teacher Expectations (TE) – Teacher Expectations refers to a teacher’s beliefs about the capability of her students to understand and access WWIGE curriculum. These are tied to the perceived cognitive, affective, and behavioral traits of students	Fun (TE - F)	Degree to which teacher expects student learning requires having fun during intervention
	Share (TE - S)	Degree to which teacher expects students must share academic experiences during intervention to learn
	Equal (TE - E)	Degree to which teacher expects students must have equal academic experiences during intervention to learn
	Enrichment (TE - En)	Degree to which teacher expects students need supplemental educational experiences to learn

Professional Experience (PE) – Professional Experience refers to overall professional experiences of teacher which impact implementation	Comfort Level (PE - CL)	Degree to which teacher believes professional experiences have prepared him/her to teach WWIGE units
	Groups (PE – G)	Degree to which teacher’s professional experiences inform decisions about grouping strategies
	Formative Assessments (PE – FA)	Degree to which teacher’s professional experiences inform decisions about when/how to use formative assessments
	Management (PE – M)	Degree to which teacher’s professional experiences inform classroom management decisions with impact implementation of WWIGE units

Table 3

Model Summaries for Poetry Unit

Parameters	Unconditional Model		Level 1 Model		Full Model	
	Parameter Estimates	SE	Parameter Estimates	SE	Parameter Estimates	SE
Intercept (γ_{00})	24.323***	.511	24.173***	.766	25.293***	.684
ITBS (γ_{10})			.019	.012	.019	.012
Teacher Adherence (γ_{01})					1.983**	.906
Residual Variance (σ^2)	19.807**	2.069	17.998***	2.147	18.118***	2.153
Intercept Variance (τ_{00})	3.588*	2.112	6.00	5.018	4.479	4.291
Deviance statistics	2008.260		1621.142		1611.142	

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 4

Model Summaries for Research Unit

Parameters	Unconditional Model		Level 1 Model		Full Model	
	Parameter Estimates	SE	Parameter Estimates	SE	Parameter Estimates	SE
Intercept (γ_{00})	25.208***	.622	25.409***	.731	27.032***	.755
ITBS (γ_{10})			.004	.008	.005	.007
Teacher Adherence (γ_{01})					3.380**	1.292
Residual Variance (σ^2)	14.669**	2.370	11.529***	2.125	11.527***	2.132
Intercept Variance (τ_{00})	7.161*	3.531	8.707*	4.198	5.905	2.744
Deviance statistics	1882.45		1420.31		1414.194	

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 5

Teaching Experience of Participants

	High Fidelity Group (N=11)		Low Fidelity Group (N=13)	
	Mean	SD	Mean	SD
Years of teaching total	14.25	8.78	20.46	9.66
Years of teaching 3 rd grade	4.44	3.33	8.58	6.29
Years of teaching gifted	3.87	1.96	6.62	6.27

Table 6

Number of Teachers with Advanced Degrees

	High Fidelity Group (N=11)	Low Fidelity Group (N=13)
Master's degree in gifted education	0	2
Master's degree in other education	6	8