

**PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Jason S. McIntosh

Entitled

THE DEPTH AND COMPLEXITY PROGRAM EVALUATION TOOL: A NEW METHOD FOR CONDUCTING INTERNAL
EVALUATIONS OF GIFTED EDUCATION PROGRAMS

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Dr. Marcia Gentry

Chair

Dr. Kristina Ayers Paul

Dr. Helen Patrick

Dr. Jenny Daugherty

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Dr. Marcia Gentry

Approved by: Dr. Richard Olenchak

Head of the Departmental Graduate Program

09/01/2015

Date

THE DEPTH AND COMPLEXITY PROGRAM EVALUATION TOOL:
A NEW METHOD FOR CONDUCTING INTERNAL EVALUATIONS
OF GIFTED EDUCATION PROGRAMS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Jason S. McIntosh

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

For my partner and parents for supporting me through this journey.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Dr. Marcia Gentry, for her guidance and support, as well as my committee members Dr. Kristina Ayers Paul, Dr. Helen Patrick, and Dr. Jenny Daugherty.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	x
ABSTRACT	xi
CHAPTER 1. INTRODUCTION	1
Statement of the Problem	1
Purpose	2
Rationale	3
Research Questions.....	5
Quantitative Research Question	6
Qualitative Research Questions	6
Mixed Methods Research Questions.....	7
Significance	7
CHAPTER 2. LITERATURE REVIEW	9
Introduction	9
Program Evaluation in Gifted Education.....	9
Trends in the Field of Program Evaluation	18
Description of Evaluative Thinking	21
Components of the DC-PET	23
Sandra Kaplan’s Depth and Complexity Model	24

	Page
Empowerment Evaluation	27
Program Evaluation Standards	37
Organizational Change and Force Field Analysis	38
Appreciative Inquiry	39
Critical Issues in the Field of Gifted Education	41
Conclusion	44
CHAPTER 3. METHODS	45
Research Design	45
Participants	46
Quantitative Data Collection and Analysis	49
Qualitative Data Collection and Analysis	52
Mixed Methods Data Analysis	58
CHAPTER 4. RESULTS	59
Quantitative Analysis	59
Factor Analysis	59
Repeated Measures ANOVA	73
Qualitative Results	75
District A Case Studies	76
District B Case Studies	101
District C Case Study	124
District D Case Study	133
District E Case Study	142
Cross-Case Analysis	148
Mixed Methods Results	167

	Page
CHAPTER 5. DISCUSSION	171
Introduction	171
Review and Discussion of Results	172
Limitations	181
Implications	182
Opportunities for Future Research	184
Conclusion	186
REFERENCES	186
APPENDICES	200
VITA	219

LIST OF TABLES

Table	Page
Table 1. Essential Components of a Gifted and Talented Program.....	14
Table 2. Attributes that Define High-Quality Program Evaluation.....	16
Table 3. Top Ranked Program Evaluation Models Recommended by Stufflebeam.....	21
Table 4. Kaplan Depth and Complexity Model Icons with Descriptions.....	26
Table 5. The Ten Principles of Empowerment Evaluation.....	30
Table 6. Description of Treatment Group.....	47
Table 7. Description of Comparison Group	48
Table 8. Quasi-Experimental Research Design	50
Table 9. Example items from Evaluative Thinking Inventory	51
Table 10. Descriptive Statistics for Pre-study Sample	62
Table 11. Original EFA Factor Loadings for Pre-Study Sample.....	63
Table 12. Final EFA Factor Loadings for Pre-Study Sample.....	64
Table 13. Descriptive Statistics for Evaluative Thinking Inventory/Prestudy.....	66
Table 14. Correlation Matrix for Pre-study Sample.....	67
Table 15. Descriptive Statistics for Post-Study Sample	69
Table 16. Descriptive Statistics for Evaluative Thinking Inventory/Poststudy.....	70
Table 17. CFA Factor Loadings for Post-Study Sample	71
Table 18. Covariance Matrix for Post-Study Sample.....	71

Table	Page
Table 19. Correlation Matrix for Post-Study Sample.....	72
Table 20. Fit Indices and RMSEA Before and After Modifications	72
Table 21. Descriptive Statistics for Matched Pairs.....	73
Table 22. ANOVA Results for Within and Between Group Variance	75
Table 23. Survey and Focus Group Response Rates for Case Study #1.....	84
Table 24. Empowerment Evaluation Principles Rating for Case Study #1	84
Table 25. Survey and Focus Group Response Rates for Case Study #2.....	91
Table 26. Empowerment Evaluation Principles Rating for Case Study #2	91
Table 27. Survey and Focus Group Response Rates for Case Study #3.....	98
Table 28. Empowerment Evaluation Principles Rating for Case Study #3.....	99
Table 29. Survey and Focus Group Response Rates for Case Study #4.....	107
Table 30. Empowerment Evaluation Principles Rating for Case Study #4	108
Table 31. Survey and Focus Group Response Rates for Case Study #5.....	115
Table 32. Empowerment Evaluation Principles Rating for Case Study #5	115
Table 33. Survey and Focus Group Response Rates for Case Study #6.....	122
Table 34. Empowerment Evaluation Principles Rating for Case Study #6	122
Table 35. Survey and Focus Group Response Rates for Case Study #7.....	130
Table 36. Empowerment Evaluation Principles Rating for Case Study #7	131
Table 37. Survey and Focus Group Response Rates for Case Study #8.....	139
Table 38. Empowerment Evaluation Principles Rating for Case Study #8	139
Table 39. Survey and Focus Group Response Rates for Case Study #9.....	146
Table 40. Empowerment Evaluation Principles Rating for Case Study #9	147

Table	Page
Table 41. Pre to Post Response Rates by Stakeholder Group for All Case Studies	150
Table 42. Themes Related to Anxieties or Concerns About Participating	152
Table 43. Average Per Person Self-Reported Number of Minutes Engaged.....	154
Table 44. Themes Related to How the DC-PET Compares to Other Methods	156
Table 45. Themes Related to What Participants Liked Best About the DC-PET.....	157
Table 46. Rating for Empowerment Evaluation Principles for all Case Studies	162
Table 47. Relevant Google Analytics Results	165
Table 48. Likelihood of Using the DC-PET Again in the Future	166
Table 49. Self-Rating of Program Evaluation Knowledge	166
Table 50. Self-Rating of Kaplan Model Knowledge	167

LIST OF FIGURES

Figure	Page
Figure 1. Web application screenshot of the DC-PET	2
Figure 2. Sample page from DC-PET workbook showing Kaplan icons	27
Figure 3. Workbook page demonstrating choices that must be made	31
Figure 4. Screenshot displaying links and resources	33
Figure 5. Workbook page describing communications plan	34
Figure 6. Screenshot illustrating inclusion of social-justice issues	36
Figure 7. Program evaluation standards cited on the workbook	37
Figure 8. Diagram of embedded quasi-experimental mixed methods design	46
Figure 9. Suggested timeline of implementation for treatment groups	55
Figure 10. Path Diagram for Evaluative Thinking Inventory	67

ABSTRACT

McIntosh, Jason, Purdue University, December 2015. The Depth and Complexity Program Evaluation Tool: A New Method of Conducting Internal Program Evaluations of Gifted Education Programs. Major Professor: Marcia Gentry.

Few program evaluation models unique to gifted education currently exist. The Depth and Complexity Program Evaluation Tool (DC-PET) is a new method for conducting an evaluation of a gifted program that combines the Kaplan Depth and Complexity Model with tools and techniques from the fields of program evaluation and organizational change. The tool was designed to assist local school district personnel in generating data for gifted program improvement by requiring a close examination of critical issues in the field (e.g., defensible differentiation, underserved populations, twice-exceptional learners). The DC-PET is meant to provide a framework for guiding those who have little or no knowledge of the evaluation process using a paper-based workbook and a computer- or tablet-based application. Gifted coordinators from five districts were asked to create one or more evaluation teams consisting of at least five stakeholders willing to pilot the DC-PET. In total, nine evaluation teams comprised of 55 participants were formed. A sample of 40 individuals from seven different school districts was used as a comparison group. Data collected from the treatment group participants included the administration of a pre and post qualitative survey, a pre and post measure of evaluative

thinking, weekly status checks, and the opportunity to participate in a focus group.

Thirty-seven participants completed pre and post assessments of their evaluation knowledge using a 4-point response scale from 1 (Novice) to 4 (Expert). Mean scores increased after 10-18 hours using the DC-PET ($M= 1.46$, $SD= 0.61$ pretest to $M= 2.19$, $SD= 0.57$ posttest). An analysis of the pre and post administration of the Evaluative Thinking Inventory (Buckley and Archibald, 2011) revealed a statistically significant interaction between the intervention and time on evaluative thinking using repeated measures ANOVA ($F(1,70) = 115.562$, $p = .027$, $\eta^2 = .068$). Further analysis of between group differences revealed no statistical difference between the treatment group and the comparison group on the pre-study version of the Evaluative Thinking Inventory, $F(1,70) = .031$, $p = .862$, meaning both groups began with about the same level of evaluative thinking. However, there was a significant difference between the treatment group and comparison group on the post-study version of the Evaluative Thinking Inventory, ($F(1,70) = 4.022$, $p = .049$, $\eta^2 = .054$). The mean evaluation team member ratings for the degree to which the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (not at all) to 4 (a lot), were 3.21 or greater. Despite early concerns regarding the time commitment and self-doubt regarding the ability to meaningfully participate, eight of the nine evaluation teams successfully completed an evaluation of a gifted program. Participants reported learning new skills and evaluation methods, as well as obtaining a greater appreciation for the importance of evaluation.

CHAPTER 1. INTRODUCTION

Statement of the Problem

Few program evaluation models unique to gifted education currently exist. This is a serious problem when one considers the importance of program evaluation to the continued existence of categorical programs such as gifted education. Insufficient expertise in evaluation methodologies by key district level leaders and a lack of commitment at the state and federal levels to gifted education (Robinson, Cotabish, Wood, & O'Tuel, 2014; Swanson & Lord, 2013) have led to program evaluation continuing to be a weak area for the field of gifted education (Warne, 2014). In fact, only 26 states include any provisions at all for gifted program accountability in their state statutes (Council of State Directors of Programs for the Gifted, 2013). This remains true despite the wide dissemination of the National Association for Gifted Children's standards, which clearly state the importance of evaluation to gifted program health (NAGC, 2010).

Many of the program evaluations of gifted programs that do take place tend to be conducted internally and lack strong, well-researched evaluation designs due to a lack of training and dedicated resources (Cotabish & Robinson, 2012; Moon, 1996; Tomlinson & Callahan, 1993; VanTassel-Baska, 2004). A logical solution might be to focus on evaluation approaches such as empowerment evaluation, a method for building

evaluation capacity within organizations to evaluate themselves (Fetterman & Wandersman, 2007). No research on the use of empowerment evaluation in the gifted education literature could be located.

Purpose

The purpose of this mixed-methods study was to pilot and examine the effectiveness of a new tool for conducting program evaluations of gifted programs entitled the Depth and Complexity Program Evaluation Tool (DC-PET, McIntosh, 2014). The DC-PET is a self-guided tool educators can use as a framework to evaluate gifted programs by following step-by-step instructions contained within an online application and a paper workbook. A screenshot displaying the home screen of the electronic app is found in Figure 1. A copy of the DC-PET workbook can be found in Appendix A.

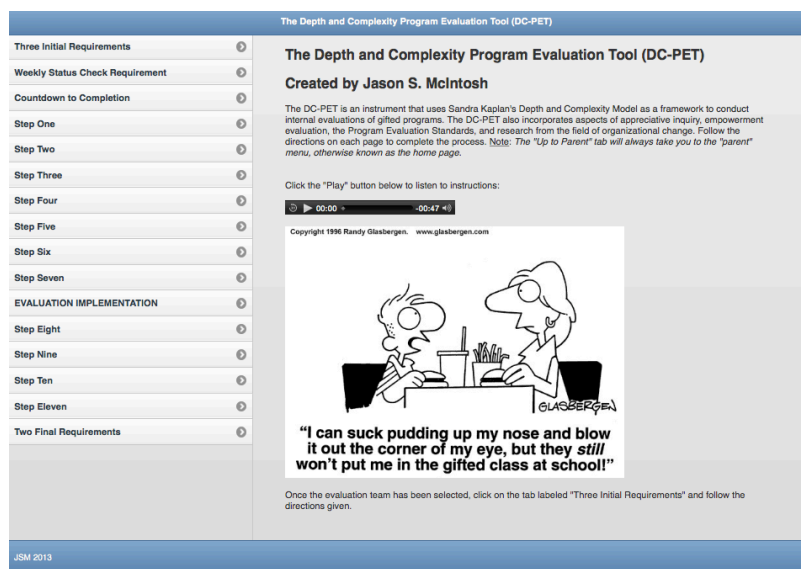


Figure 1. Web application screenshot of the Depth and Complexity Program Evaluation Tool. Cartoon © Randy Glasbergen, used with special permission from www.glasbergen.com.

An embedded quasi-experimental mixed-methods design was used. This entails the collection of qualitative data before, after, and during a quasi-experiment. Mixed-methods studies provide a wealth of evidence for studying a problem, allow researchers to study issues that could not otherwise be studied, and encourage the use of multiple perspectives (Creswell & Clark, 2007).

Rationale

In recent decades, school districts across the country have placed more importance on gathering and using valid data to make decisions. It appears that in some ways, gifted education may be behind the times. In a large-scale national survey funded by the U.S. Department of Education, less than half of the 2,000 districts surveyed reported having any strategic plan to monitor the quality of the gifted programs they offer, and a majority reported no changes to the current program would take place in the next 12-18 months (Callahan, Moon, & Oh, 2014). In a separate study, Van Tassel-Baska (2006) discovered that many of the most widely used program models in gifted education today lack credible research as to their effectiveness.

Until the financial state of gifted programs improve and priorities among administrators change, internal evaluations of gifted programs offer a low-cost alternative to no evaluation at all. This leads one to conclude that one practical solution might be to educate those conducting internal evaluations regarding effective practices in evaluation and provide easy-to-use tools and techniques. It was with this in mind that I set out to create a tool that educators with limited evaluation knowledge could use to generate rigorous, unbiased data about the programs they helm.

The DC-PET incorporates research from the fields of program evaluation, gifted education, and organizational change to create a framework stakeholders can use to evaluate a gifted program, maybe for the first time. The Pre-K-Grade 12 Gifted Programming Standards (NAGC, 2010), the Program Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (Yarbrough, Shulha, Hopson, & Caruthers, 2011), and the 10 empowerment evaluation principles developed by Fetterman and Wandersman (2007) were used as guidelines to ensure a research-based tool resulted from this effort.

Empowerment evaluation as a philosophy and process is described in detail in chapter two. In short, empowerment evaluation involves helping stakeholders build the capacity to evaluate themselves and incorporates social justice principles. The general belief is that, “The more community members participate in and control the evaluation, the more likely they are to embrace the findings and recommendations, because they own them” (Fetterman & Wandersman, 2005, p. 8).

Many practitioners in the field of gifted education are familiar with a method for teaching students to think critically known as the Kaplan Depth and Complexity Model (Kaplan, 2009). A detailed description of the Kaplan Model can be found in chapter two. The overall intent of this model is to help learners think like an expert within a specific discipline. This study represents the first time the Kaplan Depth and Complexity Model has been used to help educators think like evaluators. Applying the model in this way provided a familiar frame of reference to teachers of gifted students and assisted in making the process of conducting a program evaluation more comprehensible and less anxiety ridden.

Due to the fact many educators today feel overworked, underpaid, and underappreciated (Jacobs & Brandt, 2012), I made the decision to incorporate two techniques from the field of organizational change: (a) appreciative inquiry (Cooperrider, Whitney, & Stavros, 2008), and (b) force field analysis (Cawsey, Deszca, & Ingols, 2012). Appreciative inquiry is a method of evaluation that focuses on what is working well instead of jumping right to problem areas. Once the strengths of the program are identified, they can be capitalized upon to change aspects of the program that are not currently working at the optimal level. Celebrating accomplishments validates past efforts and helps ensure participants remain in a positive mental state as change is initiated. A force field analysis requires participants to anticipate ahead of time how the stakeholders might react to recommendations for change. Planning for these reactions in advance allows the change initiators to adapt the message and use limited resources in the most effective way possible.

Research Questions

Mixed-methods studies involve the collection of quantitative data and qualitative data at the same time. Both sets of data are analyzed separately and then merged together in an attempt to thoroughly understand an issue or phenomena (Creswell, 2014). One way to successfully merge the data is to create separate mixed-methods research questions in addition to the traditional quantitative and qualitative questions (Creswell & Clark, 2007). Quantitative, qualitative and mixed-methods research questions with summaries of how they are addressed in this study follow.

Quantitative Research Question

The quantitative research question explored in this quasi-experimental study was: *To what extent did using the DC-PET result in an increase in the participants' evaluative thinking as measured by the Evaluative Thinking Inventory* (Buckley & Archibald, 2011)? The Evaluative Thinking Inventory created by Buckley and Archibald (2011) from Cornell University's Office for Research on Evaluation (see Appendix B) was administered as a pre and post assessment. The Evaluative Thinking Inventory contains 20 questions measuring the degree to which participants: (a) pose thoughtful questions; (b) describe and illustrate thinking; (c) actively pursue deeper understanding; (d) express belief in the value of evaluation; and (e) seek alternatives. Participants were asked to select a response along a scale ranging from 1 (very frequently) to 6 (never). Exploratory factor analysis was conducted using the pre-assessment data ($N= 96$) and confirmatory factor analysis was conducted using the post-assessment data ($N= 78$), to collect reliability and validity evidence for the instrument. Upon completion, a repeated-measures ANOVA was used to discover differences between and within groups. The hypothesis related to this research question was that participants using the DC-PET will demonstrate significantly higher gains in evaluative thinking than those who do not.

Qualitative Research Questions

Four qualitative research questions were addressed in this study using focus groups, surveys, document review, and regularly scheduled status checks. The four qualitative questions were:

1. How did the DC-PET compare to previous methods of program evaluation?

2. To what extent did the Kaplan Depth and Complexity Model provide a useful framework for conducting a program evaluation?
3. To what degree did the DC-PET align with the 10 empowerment evaluation principles?
4. How did the technology component of the DC-PET affect the experience of using the tool?

Case-study methods were used to describe each treatment school followed by a cross-case analysis to look for commonalities. All qualitative data were transcribed and coded using open and axial coding as described by Creswell (2007). Member checks, a treatment group database, and a chain of evidence were compiled as described by Yin (2014).

Mixed-Methods Research Questions

All quantitative data and qualitative data were analyzed separately and then combined in order to answer two mixed-methods questions:

1. How did the qualitative results explain or expand on the quasi-experimental outcomes?
2. What modifications should be made to the DC-PET in the future based on both the quantitative results and the findings generated from the qualitative data?

Pattern matching, member checks, and the cross-case analysis were used to answer the mixed-methods questions.

Significance

This study provided useful information for determining the degree to which evaluation capacity can be built primarily through a technology tool and not through

intensive professional development of the typical sort. Second, the study demonstrated how the Kaplan Depth and Complexity Model and the 10 empowerment evaluation principles can be used as a framework for program evaluation. Third, the study resulted in the completion of eight new program evaluations in the field of gifted education and provided important information for establishing the usefulness of the DC-PET for the first time.

CHAPTER 2. LITERATURE REVIEW

Introduction

Experts in any field often observe the practices they use and the principles they believe in evolve over time due to the introduction of new ideas and technologies. Once these changes have been adopted, they are passed on and used as the new standard for excellence. Changes like these take time and do not happen overnight. A delay often exists in the time it takes for what the standard bearers profess should be happening to become the accepted practice. How program evaluations are viewed and practiced within the field of gifted education is a perfect example.

Program Evaluation in Gifted Education

Program evaluation is “necessary in creating and maintaining an exemplary, defensible gifted program” (Callahan & Hertberg-Davis, 2013, p. 6). Despite this fact, little is known about how often and to what degree local school districts conduct gifted education evaluations (Paul, 2010). It is notable that only 15 gifted program evaluation studies appeared in the literature during a 10 year period from 1990 to 2000 (Johnsen, 2000). A more recent meta-analysis could only locate 50 program evaluation studies in gifted and non-gifted journals or dissertations between 1998-2010 (Dai, Swanson, & Cheng, 2011).

A possible contributor to the problem is that few states mandate the evaluation of gifted programs (Council of State Directors of Programs for the Gifted, 2013) and only 31 include any language in high-level policy documents suggesting evaluation of gifted programs should take place at all (Paul, 2010). Van Tassel-Baska (2006) pointed to a series of interrelated issues that keep policy makers from seeing gifted education as being important to society, namely: (a) multiple conceptions of giftedness, (b) evaluation credibility issues, and (c) program models with little or no research behind them. At the local level, Moon (1996) pointed out that even if gifted program coordinators recognize the need for program evaluation they do not have the knowledge or skills to effectively and systematically implement them.

As a way to call attention to the importance of program evaluation, the National Association for Gifted Children included four indicators in the assessment standard (standard 2) related to program evaluation in the most current version of the Pre-K-Grade 12 Gifted Programming Standards (NAGC, 2010). Those standards are:

- Standard 2.5.3 - Educators assess the quantity, quality, and appropriateness of the programming and services provided for students with gifts and talents by disaggregating assessment data and yearly progress data and making the results public.
- Standard 2.6.1 - Administrators provide the necessary time and resources to implement an annual evaluation plan developed by persons with expertise in program evaluation and gifted education.
- Standard 2.6.2 – The evaluation plan is purposeful and evaluates how student-level outcomes are influenced by one or more of the following components of

gifted education programming: (a) identification, (b) curriculum, (c) instructional programming and services, (d) ongoing assessment of student learning, (e) counseling and guidance programs, (f) teachers qualifications and professional development, (g) parent/guardian and community involvement, (h) programming resources, and (i) programming design, management, and delivery.

- Standard 2.6.3 - Educators disseminate the results of the evaluation, orally and in written form, and explain how they will use the results. (p. 9)

Despite the wide dissemination of the NAGC standards, it is still unclear as to what effect they have had (Johnsen, 2014). It is also interesting to note that of the 399 studies the working group responsible for creating the standards identified as providing valid and reliable research support for their use, only six studies cited for standard two were experimental and quasi-experimental (Johnsen, 2014).

The earliest known model for evaluating gifted programs was the Diagnostic and Evaluation Scales for Differential Education for the Gifted (DESDEG) developed by Joseph Renzulli (1975). The catalyst for developing this model was the Marland Report (Marland, 1972), the first national report on gifted education to Congress. This model included: (a) an introduction to basic evaluation procedures, (b) a set of scales developed by a panel of experts in gifted education that evaluators could use as a benchmark for determining the quality of their programs, (c) a variety of data collection forms, (d) an evaluators workbook, and (e) a description of the methods for writing a summary report. Despite the development of the DESDEG, most gifted program personnel continued to use models designed for other purposes (Callahan, 1986).

A second major event in the history of program evaluation in gifted education was the founding of The National Research Center on the Gifted and Talented (NRC/GT) in 1990. One of the stated goals of the organization was, “to study the designs of gifted program evaluations in order to ascertain the effectiveness of various models and strategies” (Hunsaker & Callahan, 1993, p. 191). In an effort to fulfill this goal, Hunsaker and Callahan then critically examined 70 gifted program evaluation reports for content. Each report was coded according to 10 variables: (a) evaluation type, (b) evaluation model, (c) evaluator type, (d) data-gathering methodology, (e) data analysis technique, (f) data sources, (g) intended audiences, (h) reporting format, (i) evaluation concerns, and (j) utility information. Despite a mass mailing to 5,000 local school districts, notices in journals, and appeals at conferences, only 70 useable evaluation reports were generated (Hunsaker & Callahan, 1993). Results showed most evaluations were summative in nature, focused primarily on the concerns of program administrators, and relied heavily on data from questionnaires.

These results revealed practices different from expert evaluators in the field of program evaluation. Fleischer & Christie (2009) surveyed 1,140 professional evaluators and learned that a majority of these evaluators: (a) advocated the use of formative assessments in addition to summative measures; (b) argued for addressing the concerns of all stakeholder groups and not simply administrators; and (c) consistently triangulated data using various quantitative and qualitative measures. In addition, 91% believed planning to use the results at the beginning of the evaluation was critical. In comparison, an examination of gifted evaluation literature revealed little research reporting the use of evaluation results (Tomlinson, Bland, & Moon, 1993).

Probably the most well-known evaluation study in gifted education to date is the Arkansas Evaluation Initiative (AEI, Robinson, Cotabish, Wood, & Biggers, 2009). It was the first evaluation study to be implemented statewide. The AEI study included four components: (a) collection of program baseline data for each district; (b) participation in an Evaluation Institute to build knowledge and skills; (c) creation of templates for small, medium, and large districts; and (d) participation on a state-wide evaluation team. Randomized field studies and qualitative data showed an increase in knowledge and skills after participating in the training and completing the process.

A more recent follow-up study conducted by two of the lead authors of the AEI study focused on the effectiveness of peer coaching to increase the number of culturally, linguistically, and economically diverse (CLED) students enrolled in gifted programs throughout Arkansas. The study had disappointing results. Representatives from participating schools experienced increased evaluation knowledge just as in the original study, but the numbers of CLED students enrolled in gifted programs did not increase (Cotabish & Robinson, 2012).

A similar study conducted on a much smaller scale used the Purdue Three-Stage Model to deliver professional development designed to increase the capacity to self-evaluate a gifted program to local administrators (Moon, 1996). The sample consisted of representatives from 17 small districts who attended four training sessions over the course of a year. By the end of the year, participants were able to design quality evaluation plans, but were ultimately not successful at implementing them. Three districts carried out the evaluations as designed, but eight completed only a portion of the plan, two threw out their plans and started over, and four did not take any actions at all.

The William and Mary Eclectic Model of Gifted and Talented Evaluation took an altogether different approach to evaluation. According to the creators, the most common evaluation designs in gifted education are: (a) CIPP (Context, Input, Process, Product), (b) case study, (c) utilization-focused, (d) connoisseurship, (e) client-centered, and (f) accreditation/ certification approaches (VanTassel-Baska & Feng, 2004). The William and Mary Eclectic Model reduced these six approaches to their very essence by creating six questions to guide evaluators as they critically examine a gifted program. Although an interesting idea, the model has not been used extensively beyond the William and Mary staff.

Carter and Hamilton (1985) created a program evaluation model as well that did not gain traction. This model broke down gifted programs into nine essential components (Table 1) and listed the criteria by which each component should be judged. Carter and Hamilton provided suggested methods for data collection as well.

Table 1

Essential Components of a Gifted and Talented Program

Definition

Philosophy

Identification Procedures and Criteria

Program Goals and Objectives

Student Goals and Objectives

Curriculum

Personnel

Budget

Program Evaluation

Note. Adapted from (Carter & Hamilton, 1985)

It is much more common for researchers in the field of gifted education to advocate specific strategies or techniques for evaluating programs for gifted youth than it is for them to create a fully formed model. For example, Callahan (1983) suggested using a time-series design. This consists of delivering the same curriculum to all groups of gifted students at a grade level, but at different points during the year. The structure provides an opportunity to compare those who have and those who have not received the treatment (e.g., curriculum unit) while ensuring everyone will be exposed to the material by the end of the year. In more recent years, Callahan (2005) has created templates for assessing the rigor of school district evaluation plans and advocated for the collection of both process and outcome data to determine gifted program effectiveness. The attributes of a high-quality program evaluation as described by Callahan are shown in Table 2. These attributes were expanded upon and converted into templates that a district can use to self-assess the degree to which each attribute was fulfilled on a scale from 1 (not addressed at all) to 3 (fully addressed).

A similar approach to the templates designed by Callahan were the rubrics created by Bell (1986). A profile for a gifted program was generated by consensus, survey data collected, and the program scored using rubrics in four distinct areas: (a) philosophy, (b) teaching methods, (c) resources, and (d) student objectives. Level one on the rubric indicated no growth, whereas level four meant full implementation took place. The tool appeared in the literature only once (Bell, 1986).

Table 2

Attributes that Define High-Quality Program Evaluations

Attribute	Description
Responsiveness	Are the concerns of stakeholders considered?
Importance	Are the evaluation questions addressing important issues?
Alignment	Does the evaluation focus on the goals, objectives, and outcomes of the program?
Fairness and Impartiality	Does the evaluation give equal voice to all and are valid and reliable instruments used to collect data?
Respect for All Involved	Were the context and situational factors within the program considered? Will data actually be used?
Adequate Funding	Were the funds necessary to complete the evaluation available?
Timeliness and Relevance	Was the evaluation completed in time to make decisions about the program? Were the results and recommendations presented in a way to make the information easily understood?

Note. Adapted from (Callahan, 2005)

Articles and chapters offering general advice and recommendations related to the implementation of program evaluation in gifted education appear more frequently than specific techniques or models. A sampling of a few of these recommendations include: (a) distinguishing between advocacy and evaluation (Hunsaker, 2000); (b) using the reasonable person standard to evaluate the quality of program evaluations (Carter, 1986); and (c) using evaluation as a tool for defensibility and not as an open-ended hunting expedition (Seeley, 1986). Clinkenbeard (1996) urged districts to closely examine student motivation as a gifted program outcome, and Carter (1992) reminded districts that one of

the most important contributors to effective gifted programming is the quality of the training teachers of the gifted receive.

Callahan and Caldwell's (1997) service publication through NAGC is a notable exception. This document, *A Practitioner's Guide to Evaluating Programs for the Gifted*, articulated a six-step process that included: (a) forming evaluation questions, (b) gathering information requirements, (c) determining information sources, (d) investigating data collection strategies, (e) analyzing the data, and (f) reporting the findings. The authors gave examples of each step, provided tools practitioners could use, and aimed to remove the technical barriers and fears often experienced by those new to evaluation. The newest NAGC publication discussing the evaluation of gifted programs was released in 2012 (Speirs Neumeister & Burney). Although the authors had similar goals to those of Callahan and Caldwell, the book lacked the detail and research base of the former work (McIntosh, 2015).

The most recent line of research to emerge in gifted education program evaluation is single-subject design. Although used widely in special education in the past, it is now just beginning to be explored for use in gifted education to evaluate the effectiveness of interventions for high ability students. Single-subject research does not mean only one subject is studied, like in a case study. What it does mean, however, is each subject serves as his or her own comparison, both before and after a treatment is applied (Neuman & McCormick, 1995). To date, only two single-subject research design evaluations appear in the literature. Walsh & Kemp (2013) examined the effects of using higher order questioning on the complexity of language young gifted children use by working with a five-year-old gifted girl enrolled in a university based childcare center. The researchers

found that as the questions became more challenging, the language she used became more complex. Simonsen, Little, and Fairbanks (2010) examined the relationships among task difficulty, attention from the teacher, and off-task behavior with three gifted math students. Off-task behavior increased as teacher attention decreased, but the level of task difficulty did not have a consistent effect on off-task behavior.

Trends in the Field of Program Evaluation

There is general agreement among many evaluators that the primary function of evaluation is to provide meaningful information to stakeholders so that important decisions about programs can be made (Fleischer & Christie, 2009). The evidence-based reform movement began in the late 1990's and has continued to grow (Slavin, 2008). Ali (2001) noted that this trend resulted in most school evaluation studies morphing into accountability studies. The United States Department of Education now prioritizes randomized experiments in funding decisions over any other method (Reichardt, 2011). It is more important now than ever to use program evaluation judiciously and purposefully.

To ensure that evaluations take place in an ethical and consistent manner, the American Evaluation Association recognized and supported a set of standards for conducting evaluations developed by the Joint Committee on Standards for Educational Evaluation (Yarbrough, Shulha, Hopson, & Caruthers, 2011). These standards have been divided into five categories (i.e., utility, feasibility, propriety, accuracy, and accountability) and articulate what are considered “best practices” in program evaluation. A list of the standards can be found in Appendix C.

Studies designed to assess the perspectives of practicing evaluators confirm that the program evaluation standards accurately represent the beliefs of most evaluators

today. For example, Fleischer and Christie (2009) administered a 73 item survey to almost 900 members of the American Evaluation Association. Results showed wide spread agreement on several key issues including the importance of planning for evaluation use at the beginning of the process (91%), actively involving stakeholders in the implementation of the evaluation (98%), and communicating accurately and frequently (87%).

Not all evaluations are of equal quality and rigor. Slavin (2008) identified major issues to be aware of when reviewing the quality of individual evaluations. His suggestions included considering: (a) the sample size, (b) the duration of the evaluation, (c) if the outcome measures have inherent bias, and (d) if comparisons are made using retrospective data. Warne (2014) cautioned evaluators of gifted programs to choose instruments for measuring growth that have ample reliability and validity evidence.

Owen (2007) categorized evaluations conceptually into one of five forms: (a) proactive, (b) clarificative, (c) interactive, (d) monitoring, or (e) impact. In each conceptual form of inquiry, the evaluator serves a different function and a different approach is warranted. In a proactive evaluation the evaluator takes the role of an advisor and begins with a needs assessment along with a review of best practices before a program begins. A clarificative evaluation requires the evaluator to serve as an honest broker and begins with an assessment of evaluability once a program is initially implemented. An interactive evaluation is used for program improvement and typically involves action research or an institutional self-study. A monitoring evaluation takes place once a program is well-established and requires the use of a component analysis along with well-established performance assessments. Finally, impact evaluations are

used as a summative evaluation and use treatment and comparison groups. The forms described above are consistent with the life cycle evaluation framework created by Scheirer (2012). This framework outlines how to use evaluation during the program-development stage to test causal efficacy as the program matures, and for replication or dissemination.

A number of specific evaluation models have been developed and used over the years. Stufflebeam (2001) identified 22 in total and rank ordered them according to feasibility, accuracy, utility, and propriety. First, each of the evaluation models were put into one of the following four categories: (a) pseudoevaluations, (b) questions/methods-oriented approaches, (c) improvement/accountability oriented approaches, and (d) social agenda/advocacy approaches. Next, each model was compared to the program evaluation standards adopted by the Joint Committee on Standards for Educational Evaluation (1994). This analysis resulted in the recommended use of the nine evaluation approaches shown in Table 3. Stufflebeam (2001) included empowerment evaluation, the model strongly influencing the design of the DC-PET, under the constructivist approach, one of the chosen nine. It is notable that all four of the social agenda/advocacy approaches considered for inclusion in the recommended list were subsequently included (i.e., client-centered, deliberate democratic, constructivist, & utilization-focused). The reasons for eliminating the other 13 models included: (a) being too politically motivated, (b) lacking enough information to judge the worth of a program, (c) missing important side effects, (d) measuring only student outcomes, or (e) producing unhealthy competition (Stufflebeam, 2001).

Table 3

Top Ranked Program Evaluation Models Recommended by Stufflebeam

Decision / Accountability
Consumer Orientation
Accreditation
Utilization-focused
Client-centered / Responsive
Deliberative Democratic
Constructivist
Case Study
Outcomes Monitoring / Value-added

Note. Adapted from (Stufflebeam, 2001)

Mixed-methods designs in program evaluation continue to grow in popularity. When the American Evaluation Association created a new mixed-methods Topical Interest Group (TIG) in 2010, it rapidly became one of the largest groups in the organization (Mertens & Hesse-Biber, 2013). A second popular approach is to use logic models for evaluation purposes. A program logic model helps to create a full picture of the organization and links program activities with outcomes (W.K. Kellogg Foundation, 2004). Once the logic model has been articulated, assessments can be created to measure effectiveness or performance.

Description of Evaluative Thinking

The concept of evaluative thinking has increasingly been mentioned in evaluation journals in recent years, but consensus as to a definition has not been reached (Buckley, Archibald, Hargraves, & Trochim, 2015). The definition I used in this study is:

...critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and making informed decisions in preparation for action. (Buckley, Archibald, Hargraves, & Trochim, 2015, p. 378)

In short, evaluative thinking is a way to describe reflective practices in an organizational context (Schon, 1983) and is the “substrate that allows evaluation to grow and thrive” (Buckley, Archibald, Hargraves, & Trochim, 2015, p. 378).

The recipe for a quality evaluation requires a combination of evaluation know-how and evaluative thinking (Davidson, Howe, & Scriven, 2004). Without one the other suffers, resulting in decreased motivation, a tendency to resist change, and blind spots in perception (Buckley, Archibald, Hargraves, & Trochim, 2015). Humans are naturally inclined to belief preservation, but this can be overcome by practicing critical thinking (Lord, Ross, & Lepper, 1979). For that reason, evaluative thinking should be routinely and explicitly practiced (Buckley, Archibald, Hargraves, & Trochim, 2015).

Only two instruments have been developed to measure evaluative thinking thus far. The first is the Evaluative Thinking Assessment Tool created in 2005 by the Bruner Foundation. It was created as a result of a yearlong project known as the Evaluative Thinking in Organizations Study (ETHOS), which sought to determine how evaluative thinking is related to organizational effectiveness and how it can be fostered (Baker, Bruner, Sabo, & Cook, 2006). The tool measures the degree to which organizations possess evaluative thinking in 15 core areas (e.g., mission, strategic planning, finance, governance, leadership, communications, client development). It was developed by

modifying and combining various organizational capacity building measures found in the literature. A percentage score for each of the 15 areas is generated by dividing the number of questions marked yes by the number of possible questions for each section. Although an interesting tool, it is not appropriate for a school setting and no reliability or validity information has been published.

The second tool is the Evaluative Thinking Inventory, developed by Buckley and Archibald (2011) from Cornell University's Office for Research on Evaluation. The Evaluative Thinking Inventory contains 20 questions measuring the degree to which participants: (a) pose thoughtful questions; (b) describe and illustrate thinking; (c) actively pursue deeper understanding; (d) express belief in the value of evaluation; and (e) seek alternatives (Buckley & Archibald, 2011). Users are presented with a six-point response scale and are asked to indicate the frequency with which they take part in the five constructs listed above. The tool is practical, easy-to-use, and is very appropriate for educational settings. Little reliability and validity information has been collected, none of which appears in the literature.

Components of the DC-PET

Although some believe there is "no need to reinvent the wheel" (Seeley, 1986, p. 265), others have called for and are awaiting the development of a new program evaluation model specific to gifted education (Callahan, 2004). The DC-PET was developed by combining a critical thinking model from the field of gifted education with evaluation approaches used in the fields of program evaluation and organizational change to examine critical issues in the field of gifted education. What follows is a description of each component and how it was integrated into the tool.

Sandra Kaplan's Depth and Complexity Model

The Kaplan Depth and Complexity Model (Kaplan, 2009) is an instructional tool teachers can use to add rigor to content and foster higher-level thinking among students of all ages. The model uses 11 icons or pictures that serve as prompts for students to analyze a topic in a meaningful way. The 11 icons are shown and described in Table 4.

The first step in using the model with students is to teach the meaning of each icon. Next, the teacher provides opportunities for students to apply the icons to a topic of study together as a class. For example, the teacher might ask the students to read the *Three Little Pigs* and then identify unanswered questions they have about the story. Student responses might include such questions as, "How did the pigs learn to build houses?" or, "What would happen if the wolf had asthma?" Once students have been exposed to each icon and have had sufficient guided practice, they typically begin to use the icons with limited prompting from the teacher to create rich dialogue and deep understanding about what is being investigated. Prompting by the teacher for the students to think about the icons may come in the form of pointing to the icons displayed on the wall or giving students a graphic organizer to complete.

This study represents the first time the Depth and Complexity Model (Kaplan, 2009) has been used for a non-curricular purpose. All past applications of this model involved a teacher using the icons to prompt students to think like an expert in various fields of study. The DC-PET uses the prompts as a framework for conducting a program evaluation of a gifted program with the goal of helping the stakeholders involved in the evaluation think like expert evaluators. Using a model with which many individuals in the field of gifted education are familiar is intended to provide the scaffolding non-





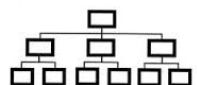
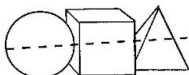

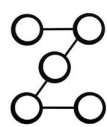

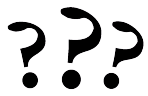

evaluators need in order to engage in the process of evaluation, possibly for the first time. A sample page indicating how the icons are utilized within the DC-PET is shown in Figure 2.

Limited research has been conducted on the Kaplan model despite the fact it has been used widely across the state of California. After an extensive search of the literature, only two empirical studies were located, both dissertations. The first study had two main research questions: (a) *How do experts utilize the prompts of Depth and Complexity in the facilitation of research and application of knowledge with their field?* and (b) *How are the prompts of Depth and Complexity relevant to the study of academic disciplines?* (Lauer, 2010, p. 37). In order to answer these questions, 10 experts in science, social science, English, and mathematics were selected and interviewed. Each of these experts met two criteria to be included. They had to have been involved in their field for at least 10 years and hold a terminal degree. The individuals came from places such as UCLA, Yale, Berkeley, Purdue, and the University of Wisconsin. The results of the study revealed that the skills the prompts of depth and complexity were designed to elicit are indeed needed and practiced in the included disciplines.

The second dissertation explored the effects of using the prompts with gifted and non-gifted students, as well as their perceptions of the model (Dodds, 2010). Two major findings were reported. First, the prompts of depth and complexity positively affected both gifted and non-gifted students alike, but gifted students had the greatest increase in understanding. Second, gifted and non-gifted students found the prompts to be interesting, useful, and challenging.


Table 4

Kaplan Depth and Complexity Model Icons with Descriptions

Name	Icon	Description
Ethics		The dilemmas or controversial issues that plague an area of study or discipline
Multiple Perspectives		The concept that there are different perspectives and that these perspectives alter the way ideas and objects are viewed and valued
Change over Time		The understanding of time as an agent of change and recognition that the passage of time changes our knowledge of things
Big Ideas		The generalizations, principles, and theories that distinguish themselves from the facts and concepts of the area or discipline under study
Rules		The natural or man-made structure or order of things that explain the phenomena within an area of study
Across the Disciplines		Integrated and interdisciplinary links in the curriculum made within, between, and among various areas of study or disciplines
Trends		The factors that influence events
Patterns		Recurring events represented by details
Language of the Disciplines		Learning the specific specialized and technological terms associated with a specific area of study
Unanswered Questions		The ambiguities and gaps of information recognized within an area or discipline under study
Details		The learning of the specific attributes, traits, and characteristics that describe a concept, theory, principle and even a fact

Note. Adapted from (Kaplan, 2009, pp.116-117)

7. When will the evaluation take place and who will be responsible for each step?

Details



Use the table below to plan when each step of the evaluation will take place and by whom. Phase one consists of how the data will be collected. Phase two, outlined in step eight, will help you analyze the data you will collect.

Program Evaluation
Standards: F4, A5, A7, U7

Phase One					
Data Collection Strategy	Date	Person Responsible	Data Collection Strategy	Date	Person Responsible

8. What patterns and trends do you see? (Phase Two)

Data can be analyzed in several ways. One useful method is to look for patterns and trends that emerge. Kaplan (2009) describes patterns as reoccurring events and trends as influences or forces that shape ideas. Use the box below to record any patterns or trends that you discover.

Patterns


Program Evaluation
Standards: U6, A7


Trends


Figure 2. Sample page from DC-PET workbook showing how icons are displayed. Arrows indicate placement of Depth and Complexity Icons.

Empowerment Evaluation

Empowerment evaluation is a research-based evaluation approach originating from the capacity building literature within the field of program evaluation (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012). The definition of empowerment evaluation used in this study was:

an evaluation approach that aims to increase the probability of achieving program success by (a) providing program stakeholders with tools for assessing the planning, implementation, and self-evaluation of their program, and (b) mainstreaming evaluation as part of the planning and management of the program or organization. (Fetterman & Wandersman, 2005, p. 28)

In short, empowerment evaluation involves helping stakeholders build the capacity to evaluate themselves, while incorporating social-justice principles. Empowerment evaluation is a close cousin to collaborative evaluation (O’Sullivan, 2004) and participatory evaluation (Cousins & Whitmore, 1998), but distinguishes itself by incorporating the 10 principles shown in Table 5. The aim of the DC-PET is to address all 10 principles to the fullest extent possible.

The first empowerment evaluation principle is the inclusion principle. This principle states that all stakeholder groups of a particular program should participate in the evaluation of the program. “The more community members participate in and control the evaluation, the more likely they are to embrace the findings and recommendations, because they own them” (Fetterman & Wandersman, 2005, p. 8). Those choosing to use the DC-PET are asked to select a diverse group of individuals (e.g., parents, teachers, students, administrators, community members) to join the evaluation team. Additionally, the team is directed to complete a chart during the data collection phase to ensure the voices of all stakeholder groups have been heard.

The second empowerment evaluation principle is the community ownership principle. This principle states that the community should be responsible for making the decisions necessary to conduct the program evaluation (Fetterman & Wandersman,

2005). Essentially, if followed, this principle gives power to the stakeholders included in the evaluation as a result of the inclusion principle to make all or most of the decisions. I incorporated this principle into the DC-PET by asking the evaluation team to: (a) collectively decide the purpose of the evaluation; (b) brainstorm and choose the evaluation questions; (c) choose or create the data collection tools; and (d) generate the recommendations and detailed goals at the conclusion of the evaluation. An example page from the DC-PET paper workbook is found in Figure 3. In many traditional evaluations, an outside evaluator or the administrator overseeing the program makes these decisions.

The third principle is the democratic participation principle. This principle is meant to ensure that the decisions made by the community are made in a democratic way (Fetterman & Wandersman, 2005). As mentioned above, users of the DC-PET are asked to determine the purpose of the evaluation, what the evaluation questions will be, and how the evaluation will be conducted. In order to ensure the decisions made are fair and represent the opinions of the majority of the group, I have included links to online project management tools like Trello (trello.com), Workflowy (workflow.com), Tasskr (tasskr.com), and Thought Boxes (thoughtbox.es). These tools allow transparent communication and collaboration between stakeholders to ensure everyone has an equal opportunity to participate. In addition, users of the DC-PET are asked to use the jigsaw technique (Aronson & Patnoe, 2011), a research-based cooperative learning strategy, to investigate issues and make decisions.

Table 5

The Ten Principles of Empowerment Evaluation


Principle	Description
#1. Inclusion	Stakeholders, including staff members, community members, funding institutions, and program participants should directly participate in decisions about an evaluation.
#2. Community Ownership	A community should make the decisions about all aspects of an evaluation, including its purpose and design; a community should decide how the results are used.
#3. Democratic Participation	Empowerment evaluations should value processes that emphasize deliberation and authentic collaboration among stakeholders; the empowerment evaluation process should be readily transparent.
#4. Community Knowledge	The tools developed for an empowerment evaluation should reflect community wisdom.
#5. Evidence-Based Strategies	Empowerment evaluation must appreciate the value of scientific evidence.
#6. Accountability	Empowerment evaluations should be conducted in ways that hold evaluators accountable to programs' administrators and to the public.
#7. Improvement	Empowerment evaluations must value improvement; evaluations should be tools to achieve improvement.
#8. Organizational-Learning	Empowerment evaluations should change organizations; cultures and influence individual thinking.
#9. Social-Justice	Empowerment evaluations should facilitate the attainment of fair allocations of resources, opportunities, and bargaining power; evaluations should contribute to the amelioration of social inequalities.
#10. Capacity-Building	Empowerment evaluations should facilitate organizations' use of data to learn and their ability to sustain their evaluation efforts

Note. Adapted from (Miller & Campbell, 2006, p. 300)

2. What is the overall purpose of the evaluation?

Four common purposes are listed below. Circle the icon that applies to this evaluation or write a different purpose on the line provided.

Contribution




- To measure the impact or effectiveness of a program

Paradox

- To uncover any weaknesses or contradictory practices

- To compare a program to the NAGC Standards or to assess compliance with state laws

Convergence



- To determine the factors that have lead to the current perceptions of a program

OTHER:

3. What unanswered questions do you have about the program you are evaluating?

Brainstorm a list below:

Unanswered Questions
???

Program Evaluation
Standards: U5, P1, P4

NOTE: The standards listed below each icon pertain to the Joint Committee on Standards for Educational Evaluation Program Evaluation Standards.

Examples:

1. Are the identification methods aligned with the services that are provided?

2 Is the curriculum rigorous enough?

Figure 3. DC-PET sample workbook page demonstrating the choices users must make in order to take ownership of the evaluation.

The fourth principle is the community knowledge principle. This principle states that stakeholders' wisdom and knowledge should be utilized throughout the evaluation (Fetterman & Wandersman, 2005). I incorporated this principle into the DC-PET at the very beginning by asking the participants to individually write down in narrative form a description of the gifted program in their own words. This valuable knowledge is referred to and then used to assist in the completion of all other tasks.

The fifth principle is the evidence-based principle. In short, this means scientific evidence and processes are valued (Fetterman & Wandersman, 2005). The DC-PET app includes numerous links to documents, tutorials, and websites describing best practices in data collection. Three examples included are the Bruner Foundation (brunerfoundation.org), the Kellogg Foundation (wkkf.org), and SAGE (us.sagepub.com). Links to previously validated instruments are also included. Three examples included are the Gifted Education Resource Institute's Instrument Repository (purduegeri.wix.com/instrument), the William and Mary Classroom Observation Scale (education.wm.edu), and the revised NAGC curriculum rubric (nagc.org). A screenshot from the online application is found in Figure 4.

The sixth principle is the accountability principle. This principle states that the people evaluating a program must be held accountable to the public and the administrators running the program (Fetterman & Wandersman, 2005). I incorporated this principle into the DC-PET by making it a requirement that a detailed communications plan be developed to clearly communicate the results of the evaluation to all stakeholders. I also provided advice on how to carefully craft the presentation of the results so that all audiences easily comprehend the content and the process undertaken to

achieve the results is made transparent. An example page from the DC-PET workbook addressing the communications plan is found in Figure 5.

The seventh and eighth principles are the improvement principle and the organizational learning principle. The improvement principle states that the goal of the evaluation should be to improve the program (Fetterman & Wandersman, 2005).

Tool Creation

Step Six

Multiple Perspectives Icon

Data Collection

Tool Creation

Up to Parent

Will you create your own tools or use previously validated instruments?

The quality of the evaluation will be determined in large part by the quality of the tools used to collect the data. A valid and reliable tool is always preferred to an untested instrument. For that reason, I have included the table below which lists previously developed and widely used tools. If a direct link is available to the tool, it has been listed under the table.

Resource	What is it?	Where do you find it?
My Classroom Activities	A validated tool for measuring elementary students' perceptions of classroom quality.	Gifted Education Resource Institute at Purdue University (see link)
Student Perceptions of Classroom Quality (SPOCQ)	A validated tool for measuring middle and high school students' perceptions of classroom quality.	Gifted Education Resource Institute at Purdue University (see link)
GERI / Purdue's Classroom Observation Tool	A classroom observation instrument used in gifted classrooms.	Gifted Education Resource Institute at Purdue University (see link)
NAGC Curriculum Rubric	A rubric developed by NAGC for the annual curriculum awards competition.	Go to the Curriculum Studies page at www.nagc.org (see link)
William and Mary Classroom Observation Scales Revised	Useful for observing gifted classrooms	Center for Gifted Education at William and Mary (see link)
Kellogg Foundation Logic Model Resource	This is an excellent resource for non-profit organizations attempting to collect outcomes oriented data.	Can be downloaded for free (see link)

[Gifted Education Resource Institute Family of Instruments](#)

[NAGC Curriculum Rubric](#)

[The William and Mary Classroom Observation Scales Revised](#)

[Kellogg Logic Model Resource](#)

Electronic surveys and questionnaires are great tools for gathering stakeholder feedback, but often appear deceptively simple to create. Quality surveys can be hard to come by. I would strongly suggest exploring the technique of survey development at the following three websites before beginning:

[General Survey Creation Principles](#)

[Rutgers List of Articles and Resources](#)

[Issues to Consider](#)

Below you will find a list of quality online survey creation tools:

[Survey Monkey](#)

[Poll Everywhere](#)

[Kwik Surveys](#)

[Zoomerang](#)

[FreeOnlineSurveys](#)

Resources for conducting a focus group can be found below:

[Toolkit for Conducting a Focus Group](#)

[Duke University's Guide to Focus Groups](#)

Tools and techniques for collecting and analyzing student data are listed below:

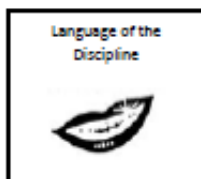
[National Association for Elementary School Principals Best Practices for Using Student Data](#)

[Student Portfolios for Measuring Growth](#)

A resource for conducting action research is listed below:

Figure 4. Screenshot of DC-PET app displaying links to previously validated instruments, data collection tools, and various resources.

11. How will you communicate your findings to the stakeholders in a clear, jargon free manner that is appropriate for each audience?



Program Evaluation
Standards: U7, P5, P6, A8

Communication plans are important for several reasons:

- Pre-change approval,
- Create need for change,
- Midstream and milestone change,
- Confirm and celebrate success

-(Cawsey, Descza, & Ingols, 2012)

In your opinion, what are the forces for and against change in your program?

Forces <u>FOR</u> Change	Forces <u>AGAINST</u> Change

Think about this as you decide what information each stakeholder needs to know and how you will present it. A few of the most common methods of presentation are:

- PowerPoint -Poster -Pamphlet -Flyer
- Podcast -Video -Website -Lecture

Use the box below to create a draft communications plan.

	Teachers	Parents	Students	Administrators
Important Information to Share				
Presentation Method/s				

FINAL NOTE: A formal written report is often required at the conclusion of an evaluation. Access to this report should be given to all stakeholders and generally includes the following:

- Title page -Table of contents -Introduction -Background
- Methodology -Results -Conclusion -Recommendations
- References -Appendices -Executive summary

Figure 5. Example DC-PET workbook page describing the development of a communications plan.

The organizational learning principle states that evaluation should change the organization for the better and influence how individuals think and operate (Fetterman & Wandersman, 2005). These two principles go hand-in-hand and were incorporated into the DC-PET in several ways. First, the users of the DC-PET must analyze the data they collected in order to determine strengths and weaknesses of the program. Next, they must generate specific goals for the future that are time bound and measurable. I refer to these as SMART goals, which is an acronym for specific, measurable, attainable, relevant and time-bound (Doran, 1981). Last, the users must communicate the results of the evaluation, including the goals for the future, to the stakeholders. In this way, knowledge is passed down to the members of the organization in hopes that it will be embraced and subsequently institutionalized.

The ninth principle is the social-justice principle. The goal of this principle is to investigate inequalities in the organization and attempt to fix them (Fetterman & Wandersman, 2005). I incorporated the social-justice principle into the DC-PET by selecting four critical issues in the field of gifted education and requiring the users of the tool to address at least one. I selected the four issues after examining the literature and reflecting on many years of experience in the field. The four social-justice issues I selected were: (a) the underrepresentation of minority students in gifted programs; (b) the provision of defensible differentiation to gifted and non-gifted students; (c) the absence of attention to the affective needs of students; and (d) the often overlooked identification of twice-exceptional students. These issues are described in detail later in this chapter. A screenshot of the DC-PET application showing how the issues are introduced can be found in Figure 6.

Equity & Excellence

Step Four

Equity & Excellence

Have you considered how your gifted program addresses the four equity and excellence issues listed below?

Stop and reflect for a moment on the four equity and excellence issues shown in the box below. A book or article citation relating to each topic has been provided as a starting point for discussion. Select one to examine more closely throughout the evaluation. Prompts will be provided periodically to focus your attention back on this issue. Record the equity issue the team selected on the DC-PET workbook (question 4).

Issue	Resource
Underrepresented populations	Ford, D. (1996). The underrepresentation of minority students in gifted education. <i>Journal of Special Education</i> , 32(1), pp. 4-14.
Twice-exceptional learners	Baum, S. (2009). <i>Talent centered model for twice exceptional students</i> . In J. Renzulli (Ed.), <i>Systems and models for developing programs for the gifted and talented</i> (pp. 17-48). Mansfield Center, CT: Creative Learning Press.
Affective needs of students	Peterson, J. S., & Moon, S. M. (2008). <i>Counseling the gifted</i> . In S. Pfeiffer (Ed.), <i>Handbook of Giftedness in Children: Psycho-Educational Theory, Research, and Best Practices</i> (pp. 225-248). New York, NY: Springer.
Defensible differentiation	Borland, J. (2009). <i>Gifted education without gifted programs for gifted students: An anti-model</i> . In J. Renzulli (Ed.), <i>Systems and models for developing programs for the gifted and talented</i> (pp. 109-118). Mansfield Center, CT: Creative Learning Press, Inc.

NOTE: I WOULD SUGGEST A JIGSAW BE USED FOR THIS STEP! Each evaluation team member chooses one issue to investigate, followed by a break of several days to allow time for information gathering and research. When the team reconvenes, all members share what they discovered and an informed decision is made by the group as to which to select.

Other sources of information on the four equity issues above can be found at:

- [National Association for Gifted Children](#)
- [Key Reports in Gifted Education](#)
- [National Research Center on the Gifted and Talented](#)
- [SENG Organization](#)
- [Critical Issues Book \(Prufrock Press\)](#)
- [Donna Ford](#)
- [Susan Baum](#)
- [Hoagies Gifted](#)

Once an equity issue has been selected, click the "Up to Parent" tab and proceed to Step Five.

Please do not forget to complete the weekly status check individually by returning to the "Weekly Status Check Requirement" tab as well.

JSM 2013


Figure 6. DC-PET app screenshot illustrating inclusion of social-justice issues.

The tenth principle is the capacity-building principle. The intent of this principle is to ensure that the individuals making up the program being evaluated and the evaluators themselves learn new skills in order to become better evaluators over time. Those using the DC-PET are introduced to new skills, techniques, and ways of thinking as they complete the 11 steps. It is unlikely that novice evaluators will complete the DC-PET without learning or refining at least one or more skills.


Program Evaluation Standards

The Joint Committee on Standards for Educational Evaluation have developed a set of standards for conducting evaluations (Yarbrough, Shulha, Hopson, & Caruthers, 2011). A shortened description of these standards can be found in Appendix C. The standards have been divided into five categories (i.e., utility, feasibility, propriety, accuracy, and accountability) and articulate what are considered “best practices” in program evaluation. These standards were used in the creation of the DC-PET and delineated for the user by noting which standard each section of the DC-PET addressed. I gave each user of the tool a copy of the standards and included codes referring to the application of each standard directly under the Depth and Complexity icon throughout. A screenshot displaying where the standards are included is found in Figure 7.

Big Ideas



Program Evaluation
Standards: P5, A1, A2,
A3, A7



Take time now to objectively compare and contrast the patterns and trends you identified within your gifted program to the NAGC standards highlighted earlier. Use the box below to generate a list of the strengths and weaknesses of your gifted program.

<i>Strengths of the Program</i>	<i>Areas for Improvement</i>

Figure 7. Example of program evaluation standards cited on the DC-PET. Arrow indicates program evaluation standards notation.

Organizational Change and Force Field Analysis

Organizational change can be defined as, “planned alterations of organizational components to improve the effectiveness of the organization” (Cawsey, Deszca, & Ingols, 2012, p. 2). Organizations typically undergo two types of change: (a) episodic change, and (b) continuous change. Episodic change involves infrequent, discontinuous, and intentional change, which may be disruptive and initiated from the managerial level (Mintzberg & Westley, 1992). Continuous change, however, is incremental, ongoing, and cumulative (Orlikowski, 1996). The sequence of events for facilitating episodic change (e.g., planned change through program evaluation) involves: (a) *unfreezing* the processes typically governing the organization; (b) *transitioning* to new ways of thinking and operating; and (c) *refreezing* to ensure fidelity of implementation (Lewin, 1951).

The idea of freezing and unfreezing implies that organizations have inertia. This physics concept explains why some organizations remain stagnant or are resistant to change and others are dynamic and adaptable (Weick & Quinn, 1999). Also consistent with the physics metaphor is the idea that forces exist for and against change, and that these forces can be overcome (Lewin, 1951). Forces against change involve any actions by an individual or group, either consciously or unconsciously, designed to prevent change from taking place (Cawsey, Deszca, & Ingols, 2012). One tool for identifying these forces is the force field analysis, which has been incorporated into the DC-PET (refer back to Figure 5). Anticipating possible arguments decision makers may use in support of or against implementing change in a program is intended to help the evaluation team prepare for and sway any arguments against positive change in the future.

Successful change agents (a) establish a sense of urgency, (b) form a change team, (c) create a vision for the change, (d) communicate the vision, (e) empower others to act, and (f) create short term wins (Kotter, 2007). Taking a bottom-up rather than top-down approach is crucial. Higgs and Rowland (2005) discovered after studying 70 different change stories that leaders who drive change through personal involvement using persuasion and influence often hinder the change they are trying to foster. This finding provides additional evidence for the use of the empowerment evaluation approach described earlier in this chapter.

Appreciative Inquiry

A second evaluation approach integral to the DC-PET is appreciative inquiry (Cooperrider, 1986). Appreciative inquiry comes from the field of organizational change and is built on five key principles: (a) the positive principle, (b) the constructionist principle, (c) the simultaneity principle, (d) the anticipatory principle, and (e) the poetic principle (Cooperrider, Whitney, & Stavros, 2008). It is different from almost every other change approach commonly used today because it focuses on what is working well within the organization instead of simply identifying problems to solve (Cooperrider & Whitney, 2005).

The positive principle states that the effort involved in identifying and celebrating strengths will enable people to transform their organizations by moving them in new, more positive, directions (Tschannen-Moran & Tschannen-Moran, 2011). In fact, proponents of appreciative inquiry believe that positive changes will begin to happen the minute a positive question is asked about the current state of the organization. This is known as the simultaneity principle. Paying attention to the positive aspects of the

working environment will naturally lead to inspiration and hope for the future (Tschannen-Moran & Tschannen-Moran, 2011). This is fittingly known as the poetic principle.

All of this is possible due to the fact organizations are created by people. Proponents of appreciative inquiry refer to this as the constructionist principle. "People do not just interpret and understand the world through their conversations and interactions with others; people thereby create the reality in which they live" (Tschannen-Moran & Tschannen-Moran, 2011, p. 423). The underlying belief here is that people hold mental models of the organization in their minds and these mental models need to be identified and examined (Norum, 2001). Once the change process begins to build steam, "positive images of the future lead to positive actions" (Norum, 2001, p. 325) through the anticipatory principle.

I incorporated appreciative inquiry into the very first step users of the DC-PET must complete. The evaluation team begins the evaluation process by discussing how their program operates when it is at its best. According to theory, beginning on a positive note and thinking about what works well will result in a more productive experience that informs any recommendations for change by providing a platform to build upon. In addition, participants are asked to collect and analyze data with the intent of determining if the mental models carried by the individuals are accurate.

A search of the literature revealed appreciative inquiry has been used on a small scale and large scale in the field of education. On a small scale, researchers in Great Britain used appreciative inquiry as part of an action research project with special needs students (Doveston & Keenaghan, 2006). On a larger scale, Tschannen-Moran and

Tschannen-Moran (2011) used appreciative inquiry across a small urban school district to increase school climate and trust.

Despite possible downsides, such as the labor-intensive nature of the approach (Austin & Harkins, 2008) and the fact collaborative efforts can lead to an increased risk of conflict (Dickerson, 2011), appreciative inquiry offers many more benefits than hurdles (Cooperrider & Whitney, 2005). Focusing on the positive aspects of schools and educational organizations will help to renew hope and dispel the myth that the system is broken. A second benefit of appreciative inquiry is that it is very flexible (Harmon & Fontaine, 2012). Simply by changing the questions asked and the stakeholders included in the process, appreciative inquiry can be used to enhance the performance of almost any organization.

Critical Issues in the Field of Gifted Education

One of the most important components of the DC-PET is the mandatory attention to one of four critical issues facing gifted educators today. Those four issues are: (a) underrepresented populations of students in gifted programs, (b) the need for defensible differentiation for all students regardless of whether they are identified as gifted, (c) the absence of attention to the affective needs of students, and (d) the often overlooked needs of twice-exceptional students. Justification for the selection of these four issues came from a review of the literature (e.g., House & Lapin, 1994; NAGC, 2010; Plucker & Callahan, 2013; Renzulli, Gubbins, McMillen, Eckert, & Little, 2009) and a reflection on my 15 years in public education teaching and coordinating gifted programs. A description of each issue has been included below.

Underrepresented populations. Unlike wealth and opportunity, giftedness should be blind to such things as race, gender, and cultural background (U.S. Department of Education, 1993). However, for decades, inequalities have existed in gifted programs favoring a bias towards White middle and upper-class students. Students from cultural, linguistic, and economically diverse backgrounds are much less likely to be nominated (McBee, 2006) and later identified (Johnsen, 2011) as gifted than their White or Asian peers. Some researchers have calculated the degree of underrepresentation to be as large as 40% (Ford, Tarek, & Gilman, 2008). There are many reasons for the problem, one of which is ineffective and inappropriate selection and identification processes (Passow & Frasier, 1996). A close examination of the representation of culturally diverse students was called for by Yoon and Gentry (2009) and could be facilitated through using the DC-PET. Doing so will help to clarify the extent of the problem for individual districts and lead to changes in policies that rectify the situation.

Defensible differentiation. It is crucial that all differentiation be defensible (Borland, 2009). The term defensible differentiation means that teachers do not reserve certain high-level tasks or skills for gifted students if all students are capable of performing them. All differentiated tasks should be respectful of the learner regardless of their ability level (Tomlinson, 1999). Although there are clear guidelines and procedures for assuring the equitable treatment of disabled students through the Individuals with Disabilities Education Act (IDEA), no such laws exist for non-disabled students. It is therefore imperative that educators closely monitor themselves in order to determine if all students, regardless of the label of gifted or not, are receiving the level of rigorous instruction they deserve. Despite a continued focus on differentiation in the last decade, it

is not commonly observed in the classroom (Tomlinson, et. al, 2003). The DC-PET may be a useful tool for carrying out the necessary examination mentioned above.

Affective needs of gifted students. Vygotsky (1986) wrote that affect and intellect are closely connected. What affects one often affects the other. The older the student becomes, the more pronounced the interaction between the two may become (Pfeiffer & Stocking, 2000). “Gifts can have both positive and negative implications” (Peterson, Assouline, & Jen, 2015, p.68) affecting social-emotional development.

Silverman noted that gifted students experience “a qualitative difference in awareness and intensity of experience” (Silverman, 2013, p.21) when compared with their typical peers. Selective schools for the gifted often include social and emotional development or prevention programs as a part of the curriculum (Eddles-Hirsch, 2012; Jones, 2011), but many traditional mixed-ability schools do not.

It is important for every school that offers academic programs for gifted students to offer social-emotional support (VanTassel-Baska, McIntosh, & Kearney, 2015).

Despite the fact educators and researchers at all levels working with all abilities of students have begun to realize the importance of the affective domain on learning and the classroom (McCoach, Gable, & Madura, 2013; Peterson, Assouline, & Jen, 2015), teachers do not often receive training in the affective needs of students and school counselors do not always follow best practices found in the literature (Wood, 2010). A focus on the affective needs of gifted students as a result of using the DC-PET may bring about much needed change.

Twice-exceptional learners. Researchers estimate that 9.1% of students who have learning disabilities, physical disabilities, emotional disabilities, or visual/auditory

disabilities are also gifted (Barnard-Brak, Johnsen, & Pond, 2009). Sadly, many of these students are never identified for gifted programs, but only receive services for their disabilities (Baum, 2009). Other students end up not receiving any services at all, gifted or special education, due to the fact that their abilities and disabilities mask each other (Brody & Mills, 1997). A close examination of the presence of twice-exceptional students through the use of the DC-PET may result in staff development and the adoption of strength-based approaches to better serve these unique learners.

Conclusion

Several major researchers in the field of gifted education have identified a general lack of expertise and skill among those responsible for conducting program evaluations at the district level. The Depth and Complexity Program Evaluation Tool (DC-PET) provides a user-friendly framework for novice evaluators and empowers district personnel to generate rigorous data that can be used for program improvement. The DC-PET combines the program evaluation standards developed by The Joint Committee on Standards for Educational Evaluation with the Kaplan Depth and Complexity Model, empowerment evaluation, and appreciative inquiry to create a program evaluation tool unique to the field of gifted education. Stephen Covey (1990) once quoted Einstein as saying, “The significant problems we face cannot be solved at the same level of thinking we were at when we created them” (p.42). The DC-PET was designed to help gifted education personnel change their perspectives, critically analyze their programs, and learn to think like evaluators.

CHAPTER 3. METHODS

Research Design

Creswell and Clark (2007) defined mixed-methods research as "a research design with philosophical assumptions as well as methods of inquiry" (p.5) that involves the collection and analysis of both quantitative and qualitative data. After examining numerous approaches, I chose an embedded quasi-experimental mixed-methods design (Creswell, 2014). This type of research design entails the collection of qualitative data before, after, and during a quasi-experiment. A diagram of the methodology can be found in Figure 8.

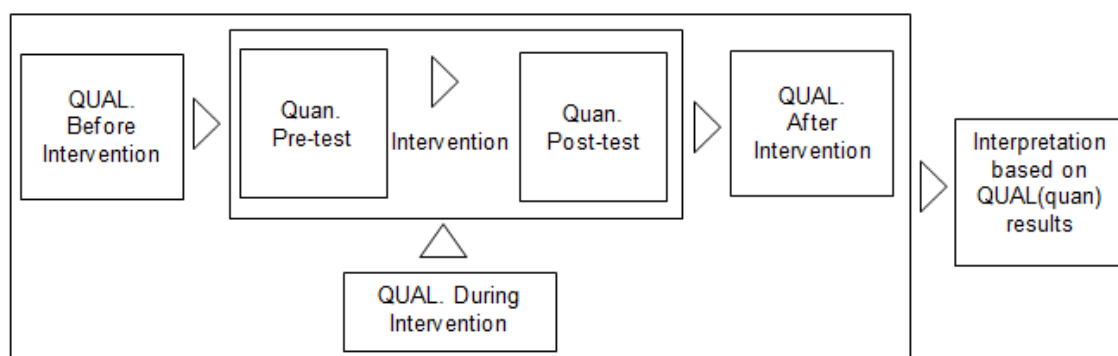


Figure 8. Diagram of an embedded quasi-experimental mixed-methods design. Adapted and reprinted with permission from Prufrock Press, Inc. (Creswell & Clark, 2007)

Participants

In addition to word of mouth and personal contacts, 25 state gifted associations and 20 State Directors of Gifted Education were contacted in order to recruit participants for the study. Nine evaluation teams from five school districts agreed to participate in the treatment group of the study. The most important qualification for participating was a willingness to dedicate time and energy over the course of six months to use the DC-PET as a way to evaluate an existing gifted program within the district. The second qualification was for each school or district to form an evaluation team of at least five stakeholders. The compilation of the stakeholders was left to the discretion of the schools, but could include parents, teachers, administrators, students, community members, or any other school staff. The total number of participants was 55.

Information including student body size, racial make-up, percentage free and reduced lunch, percentage identified as gifted, type of gifted program offered, and the number of evaluation teams along with a stakeholder classification break down for each team can be found in Table 6. Districts A, B, and C are similar in many ways and served as literal replications (Yin, 2014). All three districts have a free and reduced lunch percentage less than 30%, are suburban, have similar racial demographics, offer a true continuum of gifted program options, and are widely respected. Districts D and E are quite different, however, and helped to identify contrary patterns. Districts D and E have limited gifted programming, serve far fewer students, and consist of a majority of non-White students. A detailed description of each individual evaluation team that includes gender, years of teaching experience, and past experiences with program evaluation can be found within the case studies described in Chapter 4.

Table 6

Description of Treatment Group

	District A (Case Studies #1 - #3)	District B (Case Studies #4 - #6)	District C (Case Study #7)	District D (Case Study #8)	District E (Case Study #9)
State	AZ	AZ	AZ	AZ	MN
District Population	32,600	36,400	27,000	6,180	185
Designation	Suburb/ Public	Suburb/ Public	Suburb/ Public	Suburb/ Public	Rural/ BIE
Free and Reduced	29%	17%	22%	64%	100%
<i>Student Demographics:</i>					
-Caucasian	67%	77%	72%	28%	0%
-Black	4%	3%	3%	9%	0%
-Hispanic	24%	14%	17%	60%	0%
-Asian	4%	4%	5%	2%	0%
-Native Am.	1%	1%	2%	1%	100%
<i>GT Program Services:</i>					
-Pull-out	X	X	X		X
-Self-contain	X	X	X		
-Clustering	X	X	X	X	
-Honors	X	X	X		
-AP / IB	X	X	X		
% ID as GT	13%	6%	13%	6%	13.5%
# of Teams	3	3	1	1	1
# of Members	15	15	11	9	5
<i>Role:</i>					
-Teachers	11	8	4	3	2
-Parents	2	3	4	1	0
-Admin.	1	2	2	3	1
-Students	0	2	0	0	0
-Other	1	0	1	2	2

Note. Stats from <http://projects.propublica.org/schools/>. The category ‘other’ includes school counselor, community member, and teacher coach.

Table 7

Description of Comparison Group (N=40)

	District 1	District 2	District 3	District 4	District 5	District 6	District 7
State	AZ	AZ	AZ	AZ	GA	OH	SD
Total Population	69	250	9,920	920	82,700	3,570	483
Designation	Suburb/ Special School	Suburb/ Charter	Suburb/ Public	Rural/ Public	Suburb/ Public	Rural/ Public	Rural/ BIE
% Identified as GT	100%	16%	5%	6%	19%	5%	0%
Free and Reduced Lunch	N/A *30% on Scholar.	N/A	34%	72%	38%	9%	100%
<i>Student Demographics:</i>							
-Caucasian	84%	50%	50%	80%	35%	94%	0%
-Black	4%	5%	32%	<1%	42%	4%	0%
-Hispanic	0%	34%	10%	16%	11%	1%	0%
-Asian	12%	6%	6%	2%	9%	0%	0%
-Native Am.	0%	0%	1%	0%	0%	0%	100%
<i>GT Program Services:</i>							
-Pull-out					X	X	N/A
-Self-contain	X	X			X		
-Clustering			X	X	X	X	
-Honors			X		X		
-AP / IB					X		
<i>Number of Participants</i>	3	4	4	5	10	11	3

Note. Stats from <http://projects.propublica.org/schools>. All respondents were teachers of the gifted or coordinators of gifted programs. Suburb. = suburban.

The comparison group districts were acquired by asking those not able to dedicate the time necessary to participate as a treatment district if they would consider participating to a lesser degree in the comparison group. Word of mouth and personal contacts were used as well. The only qualification for being included in the comparison group was to be a teacher of the gifted or an administrator responsible for overseeing a gifted program. The total number of participants equaled 40. The demographics of the comparison group schools can be found in Table 7.

Quantitative Data Collection and Analysis

Research Question

The quantitative research question guiding the quasi-experimental part of this study was: *To what extent will using the DC-PET result in an increase in the participants' evaluative thinking as measured by the Evaluative Thinking Inventory?*

The hypothesis related to this research question was that participants using the DC-PET will demonstrate significantly greater gains in evaluative thinking than those who do not.

Procedures

The treatment group and comparison group participants were asked to complete the Evaluative Thinking Inventory created by Buckley and Archibald (2011) from Cornell University's Office for Research on Evaluation (see Measures section) at two different occasions to serve as a pre and post assessment. Table 8 displays the timeline used to administer the assessment. A full description of the implementation of the DC-PET required by the treatment group can be found in the qualitative research section to come.

Table 8

Quasi-Experimental Research Design

Sample One (Treatment)	O ₁	X	O ₂
Sample Two (Comparison)	O ₁		O ₂

Note. O = Collection of qualitative and quantitative data. X = exposure to the DC-PET.

Measures

The Evaluative Thinking Inventory, created by Buckley and Archibald (2011) from Cornell University's Office for Research on Evaluation, was used as a pre and post assessment in this study. Written permission was granted from the authors to use the inventory for this purpose (See Appendix D). The Evaluative Thinking Inventory contains 20 questions measuring the degree to which participants: (a) pose thoughtful questions; (b) describe and illustrate thinking; (c) actively pursue deeper understanding; (d) express belief in the value of evaluation; and (e) seek alternatives (Buckley, 2011). An example item from each construct is shown in Table 9. Participants respond to the survey items using a scale from 1 (very frequently) to 6 (never). The instrument was first presented at the American Evaluation Association conference in 2011, but has only been used on a limited scale. The pilots that have been conducted involved giving the instrument before and after a week long workshop and no statistical analyses were conducted. Due to this fact, no reliability or validity evidence was available at the time I began my study.

Table 9

Example items from Evaluative Thinking Inventory.

Construct	Example Item
Posing thoughtful questions	I pose questions about assumptions and claims made by others.
Describing & illustrating thinking	I use models and/or other diagrams to clarify my thoughts.
Active engagement in the pursuit of understanding	I discuss evaluation strategies with my colleagues.
Seeking alternatives	I consider alternative explanations for claims.
Believing in the value of evaluation	I am eager to engage in evaluation.

Data Analyses

Due to the lack of validity and reliability information on the instrument, exploratory and confirmatory factor analysis were conducted and reported using SPSS and LISREL. Exploratory factor analysis was used with the pre-study data and confirmatory factor analysis was used with the post-study data. Descriptive and inferential results were generated by comparing the treatment and comparison group's pre-study Evaluative Thinking Inventory data with the post-study data of both groups using a repeated-measures ANOVA . The effects of the independent variables (i.e., Treatment/Comparison and time) on the dependent variable (i.e., the self-reported rating of evaluative thinking) were analyzed to determine between group and within group differences.

Qualitative Data Collection and Analysis

Research Design

I used a case-study design in this study, which is a common and appropriate design for evaluation studies (Creswell, 2014; Yin, 2014). Surveys, focus groups, document review, observations, and status checks were the primary methods of collecting qualitative data. The role of the researcher was that of inquirer, peer coach, and critical friend. The following four questions were addressed:

1. How did the DC-PET compare to previous methods of program evaluation?
2. To what extent did the Kaplan Depth and Complexity Model provide a useful framework for conducting program evaluations?
3. To what degree did the DC-PET align with the 10 principles associated with empowerment evaluation?
4. How did the technology component of the DC-PET affect the experience of using the tool?

Procedures

All treatment group participants were asked to complete a nine-question survey before the study began. This survey focused on assessing the participants' previous experience with evaluation, determining the philosophical beliefs the participants have related to gifted education, and ascertaining any fears or anxieties they may have regarding participation in the study. The survey contained the following prompts:

1. Please indicate the role or roles that best describe you.
2. Describe any past experiences you have had evaluating a program.

3. Please choose the one option that best describes your current expertise with program evaluation methods. (Novice, Proficient, Advanced, Expert)
4. The purpose of evaluating a program is....
5. What concerns or anxieties do you have about participating in the DC-PET process?
6. How familiar are you with the Kaplan Depth and Complexity Model?
7. What is your definition of giftedness?
8. List three words or phrases to describe the gifted program in your district.
9. What else should the researcher know about you or your district?

I developed the survey by reflecting on the factors that might influence the participants' interaction with the DC-PET. I also pondered the data I would need to collect in order to answer my research questions. The comparison group was also given the same pre-study survey.

After completing the pre-study survey, the treatment group teams were provided with the electronic version and the paper version of the DC-PET, as well as a suggested implementation timeline (Figure 9). Next, I asked the leader of the evaluation team to arrange a meeting with all participants present. I then facilitated the first meeting to provide an orientation to the study and the DC-PET. After the question and answer period, I explained that all future meetings would take place independently of me according to the schedule they would set. I explained that I would keep track of their progress through a weekly status check they would need to complete online. The status check consisted of three questions:

1. What is going well?

2. What is not going well?
3. What questions do you have?

I monitored the status checks carefully and provided answers to the questions generated from number three shown above. In addition, I gave all participants my email address and encouraged them to contact me if questions arose. I also gave the team leads the option of inviting me back to attend a meeting in person if they felt the need to do so.

<u>Proposed DC-PET Implementation Timeline</u>		
P H A S E O N E	Introduction to the DC-PET and completion of Evaluative Thinking Inventory	One hour
	Completion of questions 1-3 (i.e. description of current program, determine purpose, brainstorm possible evaluation questions)	Two hours
	Completion of questions 4-5 (i.e. equity and excellence issue research and final evaluation question selection)	Two hours
	Completion of question 6 (i.e. development and selection of data collection instruments)	Three hours
	Completion of questions 7 (i.e. determination of evaluation schedule and role assignment)	One hour
	EVALUATION IMPLEMENTATION	Approximately 30 days
P H A S E T W O	Completion of questions 8-9 (i.e. analysis of data collected and strengths or weaknesses identified)	Three hours
	Completion of question 10 (i.e. generation of recommendations and SMART goals)	Three hours
	Completion of question 11 (i.e. force field analysis and development of communications plan)	One hour
	Completion of Evaluative Thinking Inventory and Focus Group / Reflection	One hour

Figure 9. Suggested time line for treatment group schools to implement the DC-PET.

After using the DC-PET workbook, each team member was asked to complete an online survey asking them to reflect on the experience. The survey also asked the participants to rate the degree to which they believed the DC-PET aligned with the 10 empowerment evaluation principles. The survey contained the following prompts:

1. Please choose the one option that best describes your current expertise with program evaluation methods now. (Novice, Proficient, Advanced, Expert)
2. What has lead to any changes, if any, in the level of expertise you indicated above as compared to the pre-study survey?
3. What did you like best about evaluating your program using the DC-PET?
4. What did you like the least about evaluating your program using the DC-PET?
5. How did the DC-PET compare to other methods of evaluation you have used in the past?
6. What recommendations for change would you make for the DC-PET?
7. How likely would you be to use the DC-PET in the future?
8. How familiar are you with the Kaplan Depth and Complexity Model now after using the DC-PET?
9. How likely are you to use the Kaplan Depth and Complexity Model icons with your child or students in the future?
10. Closely examine the 10 principles of empowerment evaluation listed below. To what extent do you feel the DC-PET aligned with or met the goal of each principle?

11. What else should the researcher know about the DC-PET and your experience using it?

I developed the questions after reexamining my research questions and determining what I would need to ask in order to collect enough information to answer them.

Each evaluation team member was also asked to participate in a semi-structured focus group. The focus groups were recorded and transcribed for later analysis. The questions were designed to provide an opportunity for participants to critique the tool and reflect on lessons learned. A list of recommendations for future changes was generated as well. The questions asked were:

1. How many years have you been a stakeholder at this school? Involved in education as whole?
2. What prior experiences have you had conducting research?
3. How did your group work together? Were there any conflicts? If so, how were they dealt with?
4. How long did it take for your team to complete each step on the DC-PET?
How many times did you meet in person?
5. How much time did you spend using the app v. the paper workbook?
6. What suggestions do you have for other groups using the DC-PET in the future?
7. What new skills did you learn as a result of the DC-PET?
8. What steps will you take to make sure the SMART Goals you created are implemented?
9. Did the Depth and Complexity icons help? If so, how?

10. If you went through the process again, what would you do differently?
11. Do you feel the DC-PET had too much structure, not enough, or somewhere in between?

I developed these questions by determining what was not captured in the post-study survey and what would be best asked in-person versus through a written response.

Last of all, I asked each team to provide me with the final evaluation reports generated as a result of this study and copies of completed DC-PET workbooks. Follow-up questions were asked of the participants after reading the final evaluation reports in order to clarify claims made or procedures followed. Google Analytics, a free service offered by Google that tracks and records website traffic, was used to determine the number of times the online version of the DC-PET was accessed and the average length of each session.

Data analyses

Careful records were kept by creating a case study database and chain of evidence record. This database included a separate section for each evaluation team containing all evidence collected including completed surveys, focus group transcripts, status checks, data collection instruments created by the teams, and final evaluation reports.

After all transcripts were transcribed, they were read and coded using open and axial coding as described by Creswell (2007). Patterns were identified and categories were determined. Each evaluation team was described in a separate case study before aggregating the data across groups. This is known as a within-case analysis (Creswell, 2007). Next, all case studies were combined to conduct a cross-case analysis.

Triangulation of data was achieved by checking to see that multiple participants both within and across cases observed important findings. I also conducted a member check at the conclusion of the analysis to determine the degree to which the evaluation team members agreed with the first draft of each case study. A thick description of the data was included in the final results section and close attention was given to any contrary evidence. Potential alternative explanations for results were explored.

Mixed-Methods Data Analysis

The results of the quantitative and qualitative analyses described above were then combined to answer the following two questions:

1. How did the qualitative results explain or expand on the quasi-experimental outcomes?
2. What modifications should be made to the DC-PET based on the quantitative results and the themes generated from the qualitative data?

Pattern matching and the cross-case analysis were used to answer the mixed-methods questions.

It is important for any researcher to recognize that all writing is positioned within an individual's worldview and past experiences (Creswell, 2007). Subtexts, biases, and personal subjective interpretations can exist. I made every attempt to bracket my beliefs and experiences in order to see the process through the eyes and ears of the participants. The member checks previously described were used to help with this goal.

CHAPTER 4. RESULTS

This chapter is divided into three sections. Section one contains the results of the quantitative procedures used in this study (i.e., exploratory factor analysis, confirmatory factor analysis, repeated measures ANOVA). Section two contains the qualitative results in the form of nine individual case studies, one for each evaluation team, and one cross-case analysis. Section three contains the results of introducing the quantitative data to the qualitative data in order to explore the mixed-methods research questions described in chapter three.

Quantitative Analysis

Three major quantitative techniques were used to answer the research question, *To what extent will using the DC-PET result in an increase in the participants' evaluative thinking as measured by the Evaluative Thinking Inventory?* Exploratory and confirmatory factor analysis using SPSS and LISREL were used to collect validity evidence of the Evaluative Thinking Inventory and a repeated measures ANOVA was used to examine patterns of change from the first administration of the Evaluative Thinking Inventory to the second administration.

Exploratory Factor Analysis

SPSS was used to conduct an exploratory factor analysis on 96 Evaluative Thinking Inventories completed at the beginning of the study by both the treatment and comparison groups. The Evaluative Thinking Inventory contains 20 questions measuring

the degree to which participants: (a) pose thoughtful questions; (b) describe and illustrate thinking; (c) actively pursue deeper understanding; (d) express belief in the value of evaluation; and (e) seek alternatives (Buckley & Archibald, 2011). Participants must choose a response from 1 (very frequently) to 6 (never).

The first step in conducting the exploratory factor analysis was to clean the data. This involved filling in two pieces of missing data with the mean for the group and removing one multivariate outlier. The final descriptive statistics for each question on the Evaluative Thinking Inventory can be found in Table 10.

Originally, the inventory was designed with five factors as described above. Four questions were created for each factor based on theory. When a five-factor model was forced using the principal axis extraction method and a direct oblimin factor rotation, the items did not correlate as expected. The initial factor loadings can be found in Table 11. In fact, there were numerous cross-loadings, questions that did not load on the hypothesized factor, and questions that did not load on any factor. Direct oblimin rotation was used due to the correlated nature of the factors (McCoach, Gable, & Madura, 2013).

After purposefully exploring the data to look for patterns and using trial and error to delete questions one at a time, a noticeable trend emerged. Certain items within the *posing thoughtful questions* factor and the *seeking alternatives* factor were loading on the same factor instead of separately. Also, items within the *expressing belief in the value of evaluation* factor and the *actively pursuing deeper understanding* factor were loading on the same factor instead of separately. The ties between these constructs made logical sense and so a three-factor model was run. The new factors were renamed: (a) *describing*

and illustrating thinking; (b) posing thoughtful questions and seeking alternatives; and (c) believing in and practicing evaluation.

The new 14 item three factor model, with factor loadings <0.4 suppressed, is shown in Table 12. The path diagram is shown in Figure 10 and alpha reliabilities with descriptive statistics by factor are shown in Table 13. The alpha-reliability estimates of the data for internal consistency were 0.815 for factor one, 0.798 for factor two, and 0.888 for factor three. A Cronbach Alpha greater than 0.70 is considered adequate for an affective measure (McCoach, Gable, & Madura, 2013). It is also notable that a majority of individuals chose responses 1-3 when responding to each statement. A lower rating corresponds to a higher frequency of participating in that action. Together, the three factors accounted for 49.68% of the variance. The correlation matrix for the final model is shown in Table 14.

One issue still remained. Factor three now only had two questions that are 0.4 or greater. Standard practice dictates that a factor must have a minimum of three questions (Brown, 2006). The results of the exploratory factor analysis were presented to the creators of the inventory, Buckley and Archibald, through a teleconference and an agreement was made to create and pilot additional questions for the *describing and illustrating thinking* factor at a later date.

Table 10

Descriptive Statistics for Pre-Study Sample / Treatment & Comparison (n=96)

	<i>M</i>	<i>SD</i>	Variance	Skewness	Kurtosis
Item 1	2.03	.839	.704	.923	1.249
Item 2	2.17	.660	.435	-.191	-.702
Item 3	1.96	.695	.482	.248	-.264
Item 4	2.40	.968	.936	.551	.797
Item 5	1.55	.663	.439	1.023	.884
Item 6	2.47	.894	.799	.457	.615
Item 7	2.21	.710	.504	.759	1.960
Item 8	2.00	.632	.400	.000	-.434
Item 9	1.79	.780	.609	.521	-.739
Item 10	1.72	.721	.520	.482	-.947
Item 11	1.50	.616	.379	.830	-.286
Item 12	2.35	.833	.694	.254	.148
Item 13	1.99	.788	.621	.414	-.313
Item 14	2.08	.763	.582	.148	-.597
Item 15	1.80	.803	.645	.999	1.559
Item 16	2.58	.959	.919	.709	1.029
Item 17	1.96	.664	.440	.045	-.682
Item 18	2.41	1.052	1.107	.724	.589
Item 19	1.68	.673	.453	.490	-.743
Item 20	2.35	.846	.715	.311	.113

Table 11

Original EFA Factor Loadings for Pre-Study Sample / Treatment & Comparison (n=96)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
20. I enjoy discussing evaluation strategies with colleagues.					0.690
6. I discuss evaluation strategies with my colleagues.	0.606				0.315
16. I try to convince others that evaluation is important.	0.565				0.365
12. I articulate the relationship between my evaluation work and my intended claims.	0.740				
7. I articulate the logical justification of my evaluation strategy.	0.460				
10. I seek evidence for claims and hypotheses.		0.974			
9. I am wary of claims made by others without evidence to back them up.		0.740			
11. I am interested in understanding the logic behind things.		0.576			
14. I pose questions about assumptions and claims made by others.	0.453		0.356		
13. I reflect on assumptions and claims I make myself.	0.530			0.425	
17. I offer evidence for claims that I make.			0.313		
5. I take time to reflect about the way I do my work.					
4. I use models and/or other diagrams to clarify my thoughts.				0.998	
18. I use models and/or other diagrams to communicate my thinking to others.				0.785	
2. I am eager to engage in evaluation.	0.483				0.637
1. I describe my thinking to others.				0.560	
19. I believe evaluation is a valuable endeavor.					0.533
3. I suggest alternative explanations and hypotheses.					
8. I consider alternative explanations for claims.			0.344	0.316	
15. I willingly make changes to the way I do my work.					

Note. Items loading <0.3 were suppressed.

Table 12

Final EFA Factor Loadings for Pre-Study Sample / Treatment and Comparison (n=96)

	Factor 1	Factor 2	Factor 3
20. I enjoy discussing evaluation strategies with colleagues.	0.879		
2. I am eager to engage in evaluation.	0.681		
16. I try to convince others that evaluation is important.	0.606		
19. I believe evaluation is a valuable endeavor.	0.569		
6. I discuss evaluation strategies with my colleagues.	0.563		
14. I pose questions about assumptions and claims made by others.		0.682	
17. I offer evidence for claims that I make.		0.648	
8. I consider alternative explanations for claims		0.606	
9. I am wary of claims made by others without evidence to back them up.		0.549	
13. I reflect on assumptions and claims I make myself.		0.543	
3. I suggest alternative explanations and hypotheses.		0.515	
5. I take time to reflect about the way I do my work.		0.407	
4. I use models and/or other diagrams to clarify my thoughts.			0.924
18. I use models and/or other diagrams to communicate my thinking to others.			0.840

Note. Items loading <0.4 were suppressed.

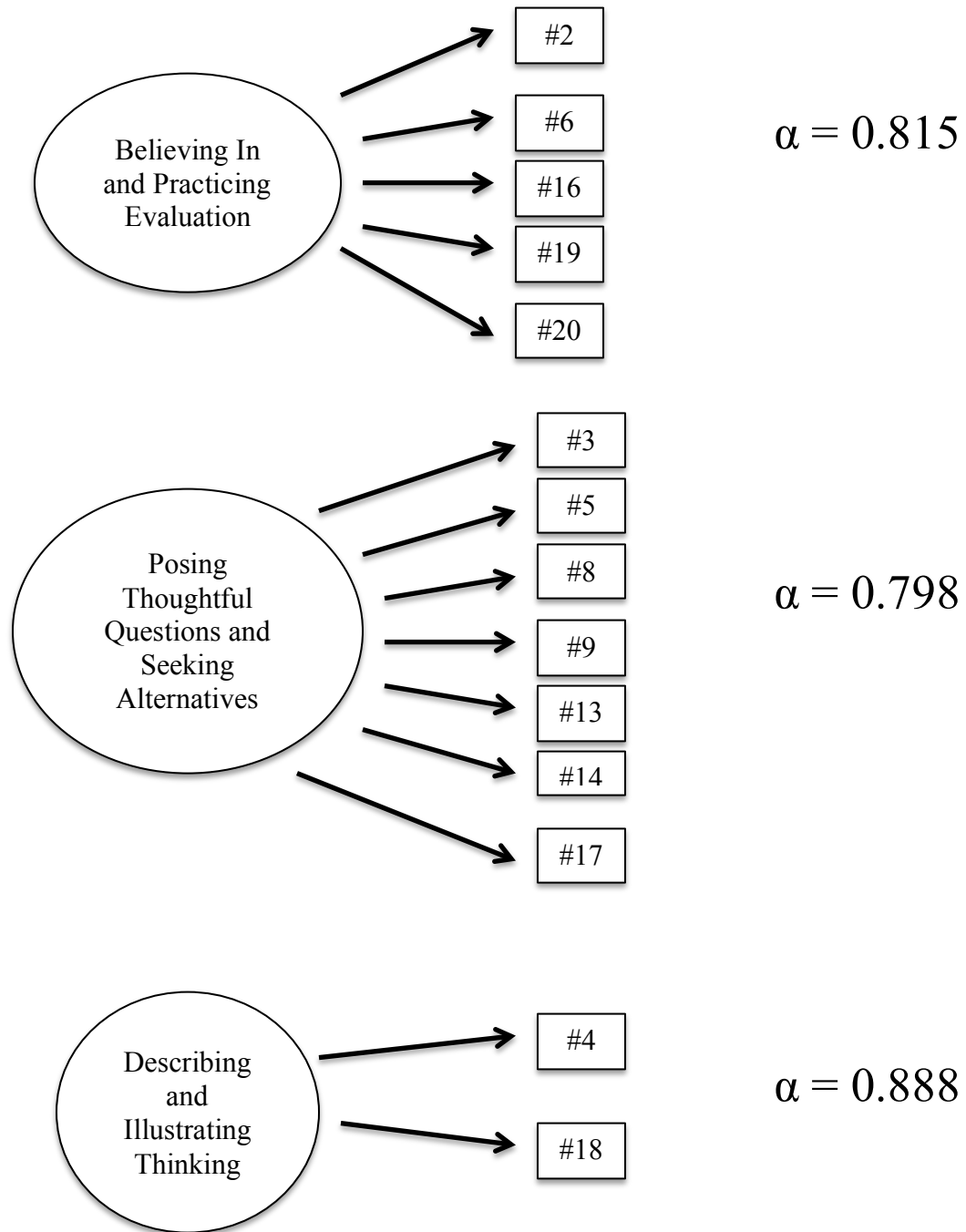


Figure 10. Path diagram for the Evaluative Thinking Inventory after exploratory factor analysis.

Table 13

*Evaluative Thinking Inventory Descriptive Statistics by Factor for Pre-Study Sample**(n=96)*

Factor	Item	Response Percentage						<i>M</i>	<i>SD</i>	<i>r</i> with corrected item total	Alpha if item removed	Alpha
		1	2	3	4	5	6					
Believing in and practicing evaluation	2	15	54	31	0	0	0	2.17	0.66	.55	.80	.82
	6	12	40	40	5	3	0	2.47	0.89	.54	.80	
	16	9	42	34	12	2	1	2.58	0.96	.66	.77	
	19	44	45	11	0	0	0	1.68	0.67	.60	.78	
	20	15	44	34	6	1	0	2.35	0.85	.72	.74	
Posing thoughtful questions and seeking alternatives	3	25	55	19	1	0	0	1.96	0.69	.55	.77	.80
	5	53	40	6	1	0	0	1.55	0.66	.50	.78	
	8	20	60	20	0	0	0	2.00	0.63	.55	.77	
	9	42	38	19	1	0	0	1.79	0.78	.40	.80	
	13	28	48	21	3	0	0	1.99	0.79	.56	.77	
	14	23	48	27	2	0	0	2.08	0.76	.60	.76	
	17	24	56	20	0	0	0	1.96	0.66	.59	.76	
Describing and illustrating thinking	4	18	39	32	10	0	1	2.40	0.97	.80	.85	.89
	18	15	44	34	6	1	0	2.41	1.05	.80	.84	

Note. Scale from 1 – 6 (Very Frequently to Never)

Correlation Matrix for Pre-Study Sample / Treatment & Comparison (n=96)

[illegible]

Confirmatory Factor Analysis

LISREL code was generated from the results of the exploratory factor analysis and used to test the Evaluative Thinking Inventories taken by both the treatment and comparison groups at the conclusion of the study ($n=78$). The original 20-question version of the inventory was given. However, only items loading on factors one and two identified in the exploratory factor analysis were analyzed. Factor three was ignored due to the fact that it had only two questions (Brown, 2006). Items 2, 6, 16, 19, and 20 loaded onto the *believing in and practicing evaluation factor*. Items 3, 5, 8, 9, 13, 14, and 17 loaded onto the *posing thoughtful questions and seeking alternatives factor*. The descriptive statistics for the sample are located in Table 15 and alpha reliabilities with descriptive statistics by factor are located in Table 16. Item factor loadings are shown in Table 17. The covariance matrix among factors is shown in Table 18 and the correlation matrix among items is shown in Table 19.

The χ^2 value was significant at 91.788 ($p=0.0008$). The Goodness of Fit Index (GFI) was 0.837, which is lower than the optimal 0.90 or greater. However, the Comparative Fit Index (CFI) was 0.901 and the Incremental Fit Index (IFI) was 0.905. The RMSEA was 0.097, which is greater than the recommended 0.08 (Brown, 2006). When two correlated errors were freed based on the maximum modification index, the χ^2 was no longer significant ($p=0.06$). The GFI became 0.886, the CFI became 0.958, and the IFI became 0.960. The RMSEA also fell below the 0.08 threshold to 0.064. The freed correlated errors were between questions 3 and 8 followed by questions 13 and 14. A full description of the changes due to the two modifications can be found in Table 20. The CFA results are strong and indicate a good fit for the data.

Table 15

Descriptive Statistics for Post-Study Sample / Treatment & Comparison (n=78)

	Mean	SD	Variance	Skewness	Kurtosis
Item 2	1.92	.864	.747	.770	.687
Item 3	2.00	.773	.597	.866	1.994
Item 5	1.46	.574	.330	.787	-.360
Item 6	2.21	.858	.737	.473	.376
Item 8	2.03	.772	.597	.303	-.410
Item 9	1.77	.852	.725	.726	-.532
Item 13	1.84	.779	.607	.624	-.074
Item 14	2.05	.771	.595	.609	.411
Item 16	2.29	.884	.782	.188	-.101
Item 17	1.78	.617	.380	.168	-.496
Item 19	1.63	.626	.392	.798	1.285
Item 20	1.95	.804	.647	.402	-.555

Table 16

Evaluative Thinking Inventory Descriptive Statistics by Factor for Post-Study (n=78)

Factor	Item	Response Percentage						<i>M</i>	<i>SD</i>	<i>r</i> with corrected item total	Alpha if item removed	Alpha
		1	2	3	4	5	6					
Believing in and practicing evaluation	2	36	40	22	1	1	0	1.94	0.86	.61	.83	.85
	6	20	45	30	4	1	0	2.22	0.85	.62	.83	
	16	21	36	38	4	1	0	2.31	0.88	.66	.82	
	19	44	51	4	1	0	0	1.64	0.63	.66	.83	
	20	32	43	22	3	0	0	1.96	0.80	.80	.78	
Posing thoughtful questions and seeking alternatives	3	25	55	18	1	1	0	2.01	0.77	.39	.84	.83
	5	58	38	0	0	0	0	1.45	0.58	.58	.81	
	8	25	49	23	3	0	0	2.03	0.78	.69	.79	
	9	47	31	19	3	0	0	1.77	0.86	.50	.83	
	13	36	45	16	3	0	0	1.84	0.78	.68	.79	
	14	22	57	16	5	0	0	2.05	0.78	.64	.80	
	17	32	58	10	0	0	0	1.78	0.62	.67	.80	

Note. Scale from 1 (Very Frequently) to 6 (Never)

Table 17

CFA Factor Loadings for Post-Study Sample / Treatment and Comparison (n=78)

	Factor 1	Factor 2
2. I am eager to engage in evaluation.	0.631	
6. I discuss evaluation strategies with my colleagues.	0.708	
16. I try to convince others that evaluation is important.	0.710	
19. I believe evaluation is a valuable endeavor.	0.711	
20. I enjoy discussing evaluation strategies with colleagues.	0.910	
3. I suggest alternative explanations and hypotheses.		0.425
5. I take time to reflect about the way I do my work.		0.614
8. I consider alternative explanations for claims		0.717
9. I am wary of claims made by others without evidence to back them up.		0.570
13. I reflect on assumptions and claims I make myself.		0.778
14. I pose questions about assumptions and claims made by others.		0.751
17. I offer evidence for claims that I make.		0.724

Table 18

Covariance Matrix for Post-Study Sample

Factor	Covariance	
One	1.000	
Two	0.569	1.000

Table 19

Correlation Matrix for Post-Study Sample / Treatment & Comparison (n=78)

Item 2	1.000												
Item 6	.424	1.000											
Item 16	.489	.484	1.00										
Item 19	.546	.434	.529	1.000									
Item 20	.555	.674	.642	.632	1.000								
Item 3	.330	.078	.133	.349	.293	1.000							
Item 5	.229	.253	.266	.303	.277	.263	1.000						
Item 8	.334	.286	.255	.396	.441	.566	.383	1.000					
Item 9	.099	.137	.161	.348	.400	.138	.433	.404	1.000				
Item 13	.122	.350	.226	.287	.390	.245	.484	.506	.418	1.000			
Item 14	.103	.396	.320	.282	.318	.218	.444	.543	.354	.688	1.000		
Item 17	.212	.184	.310	.359	.396	.300	.471	.502	.496	.554	.515	1.000	

Table 20

Fit Indices and RMSEA Before and After Modifications for Post-Study Sample

	Before Modification	After Modification #1	After Modification #2
GFI	0.837	0.868	0.886
CFI	0.901	0.944	0.958
IFI	0.905	0.945	0.960
RMSEA	0.097	0.074	0.064
χ^2	91.788	74.193	67.471

Repeated Measures ANOVA

A repeated measures ANOVA was chosen to analyze the pre and post Evaluative Thinking Inventory data collected from the treatment and comparison groups. There were no outliers, as assessed by inspection of a boxplot for values greater than 1.5 box lengths from the edge of the box. The data was normally distributed for all interventions and time points, as assessed by Shapiro-Wilk's test ($p>.05$). Homogeneity of variance existed, as assessed by Levene's test of homogeneity of variance ($p>.05$). Homogeneity of covariances also existed, as assessed by Box's test of equality of covariance matrices ($p>.001$). Descriptive statistics for the data can be found in Table 21. The design of the scale translates to a smaller score indicating a person or group thinks more like an evaluator than a person or group with a higher score.

Table 21

Descriptive Statistics for Matched Pairs Treatment and Comparison Group for Evaluative Thinking Inventory

		<i>N</i>	<i>M (SD)</i>	<i>SE</i>	<i>95% CI</i>
Pre-Study	Treatment	36	40.97 (7.86)	1.347	(38.29, 43.66)
	Comparison	36	41.31 (8.30)	1.347	(38.62, 43.99)
	Total	72	41.14 (8.03)		
Post-Study	Treatment	36	35.61 (8.99)	1.381	(32.86, 38.37)
	Comparison	36	39.53 (7.50)	1.381	(36.77, 42.28)
	Total	72	37.57 (8.46)		

There was a statistically significant interaction between the intervention and time on evaluative thinking, $F(1,70) = 115.562, p = .027, \eta^2 = .068$. Further analysis of between group differences revealed no statistical difference between the treatment group and the comparison group on the pre-study version of the Evaluative Thinking Inventory, $F(1,70) = .031, p = .862$, meaning both groups began with about the same level of evaluative thinking. However, there was a statistically significant difference between the treatment group and comparison group on the post-study version of the Evaluative Thinking Inventory, $F(1,70) = 4.022, p = .049, \eta^2 = .054$, which is 5.4% of the variance explained. Partial eta squared (η^2) is the “ratio of variance accounted for by an effect and that effect plus its associated error variance within an ANOVA study” (Brown, 2008, p.40). This provided evidence that using the DC-PET increased the frequency with which those participating in the treatment group thought like an evaluator. The results of the between group analysis can be found in Table 22.

Analysis of within group differences revealed evaluative thinking was statistically different between pre and post administration of the Evaluative Thinking Inventory for the treatment group, $F(1,35) = 15.635, p < .0005, \eta^2$ of .309. This indicates 30.9% of the within group difference can be explained by time (taking the pre or post survey). The comparison group also had a statistical difference between pre and post administrations of the Evaluative Thinking Inventory, $F(1,35) = 4.566, p = 0.040, \eta^2 = 0.115$, which represents 11.5% of the within group variance explained by time (taking the pre or post survey). The results of the within group analysis can also be found in Table 22.

Table 22

ANOVA Results for Within and Between Group Variance

	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Between Subjects						
Pre-Study	2.000	1	2.000	0.031	0.862	<0.001
Error	4574.611	70	65.352			
Post-Study	276.125	1	276.125	4.022	0.049	0.054
Error	4805.528	70	68.650			
Within Subjects						
Treatment	517.347	1	517.347	15.635	0.000	0.309
Error	1158.153	35	33.090			
Comparison	56.889	1	56.889	4.566	0.040	0.115
Group Error	436.111	35	12.460			

Qualitative Results

The qualitative results are presented in the form of nine case studies and a cross-case analysis. Case studies 1-3 describe the evaluations conducted in District A. Case studies 4-6 describe the evaluations conducted in District B. Case studies 7-9 describe evaluations conducted in three separate districts (i.e., District C, District D, District E). A successful evaluation using the DC-PET is defined as one in which the participants have completed steps 1-8 or greater, and plan to finish any remaining steps the following semester. In other words, a successful team has generated an evaluation plan, created or selected the tools necessary to collect the data, collected data using those tools, and

begun the process of analyzing the data. The communications plan and the creation of SMART goals could be scheduled for a later date after the conclusion of this study.

District A Case Studies

District A has one of the largest, well-funded gifted programs in the state of Arizona. Currently, gifted services include the typical options (e.g., self-contained gifted classrooms, gifted cluster groups, International Baccalaureate programs) and more creative options (e.g., digital academy for advanced placement scholars, school-within-a-school science and engineering research center, services for Pre-K gifted students). The district is comprised of 61 schools and received an overall grade of 'B' on the 2013/2014 state report card generated by the Arizona Department of Education.

Three separate evaluation teams were formed to examine the effectiveness of three different service models used in the district: (a) the gifted cluster model, (b) content replacement, and (c) self-contained gifted classes. Unlike the other districts involved in this study, each of the evaluation teams were compensated by District A for the time they spent working on the evaluation. I facilitated the first meeting for each team and the remaining sessions were conducted independently.

Case study #1. Case study one describes the gifted cluster classroom evaluation team within District A. Cluster classrooms were created in grades 1-4 at eight elementary schools within the district. The practice of cluster grouping, as described by the district website, is placing gifted students in mixed-ability classrooms at each grade level with a teacher that has had training in understanding, planning for, and instructing gifted students.

Participants. The gifted cluster evaluation team consisted of three gifted cluster classroom teachers, a teacher coach, and a parent of a gifted student enrolled in the program. All evaluation team members were White women. Teaching experience ranged from 5 to 18 years. One teacher taught fourth grade, one taught third grade, and one taught first grade during the 2014/2015 school year. Three of the team members reported no past experiences evaluating a program. The remaining two members discussed informally evaluating a classroom and piloting a math curriculum.

Pre-study survey results. An analysis of the pre-study surveys completed by these five individuals revealed a general consensus among the team members that the gifted program in their district was strong, challenging, and innovative. All of the adjectives used to describe the program in question eight of the survey (i.e., *List three words or phrases to describe the gifted program in your district*) were positive. Other adjectives listed included differentiated, growing, independent, organized, and student-centered.

The mean self-reported rating for existing program evaluation expertise was 1.4 (0.89) on a scale from 1 (Novice) to 4 (Expert). This placed the team at the novice level. All team members believed that the purpose of evaluation is to examine the effectiveness of a program in order to improve in the future. When asked to list any concerns or anxieties they may have about participating, two team members wrote none. A second pair mentioned the time it would require, and one mentioned a feeling of anxiety about discovering personal weaknesses.

The mean self-reported rating of pre-existing knowledge of the Kaplan Depth and Complexity Model was 2.6 (0.55) on a scale from 1 (I've never heard of it) to 4 (I

use it all the time). This placed the team between “I know what it is” and “I’ve used it once or twice”.

Experience while using the DC-PET. The team members met together once a month between September and April to complete the tool. Google Hangouts, an online collaboration tool, was used for two meetings and five were conducted in person. Overall, the group reported spending 13-15 hours on the process. Although there are 11 steps, the team only completed 10 on their own due to a perceived lack of authority from the Director of Gifted Education to create SMART goals and a communications plan. The Director of Gifted Education later completed the final tasks.

The team successfully completed 16 online status checks while using the tool. The content of these status checks included: (a) reporting that the team was collaborating well, (b) a concern that an optional reading about underrepresented populations included in step four of the tool (i.e., Ford, 1998) was outdated, and (c) a question related to ensuring the survey instrument they created was valid and reliable. Team members reported no questions or concerns a total of 13 times. In response to these status checks, I sent the group a new chapter focusing on underrepresented populations entitled *Being Gifted and Adolescent: Issues and Needs of Diverse Students* (Worrell, 2015) and offered to provide feedback on the survey they created. I received no requests to attend additional meetings in person after the initial meeting.

The team reported using the online version of the DC-PET more heavily in the beginning due to the videos and links it contained, but the paper workbook more heavily during the meetings and towards the end of the process. The consensus of the group was that both were needed and served different functions.

Description of the evaluation conducted by the team. District A's cluster grouping evaluation team decided the purpose of the evaluation would be to measure the effectiveness of the cluster model used within the district. The team generated numerous possible unanswered questions, but chose to explore:

- Is differentiation being implemented in gifted cluster classrooms?
- Are there identifiable factors that affect differentiation in gifted cluster classrooms?
- To what degree is the differentiation taking place in the cluster grouping classrooms defensible?

The first two questions came from the team's brainstorming efforts and the third question came from choosing an equity and excellence issue as described in step four.

In order to answer their evaluation questions, four separate surveys were carefully crafted. The corresponding surveys were sent to teachers, parents, students, and administrators. The number of respondents from each stakeholder group included 5 administrators, 22 gifted cluster teachers, 29 gifted students, and 67 parents.

The administrator surveys revealed that all respondents understood the importance of classroom composition to the implementation of the cluster model and saw evidence of differentiation in the classrooms they observe. Furthermore, four out of five administrators reported differentiation taking place in the cluster classrooms on a routine basis. The evidence cited by the administrators included choice boards, digital learning, ability grouping, projects, and self-guided learning.

The teacher survey contained questions designed to determine what type of differentiated activities they offered to their students. Results showed: (a) 68% reported

using curriculum compacting, (b) 49% reported using extension menus, (c) 82% reported providing acceleration opportunities, and (d) 77% reported using tiered lesson planning. The cluster teachers were also asked to list any challenges they face on a daily basis. Challenges listed included: (a) a need for more time to plan and create; (b) a desire for additional materials; and (c) a belief that cluster teachers have too wide a range of student abilities in their classrooms.

Questions for the student survey came from Gentry and Gable's (2000) *My Class Activities* instrument. Instead of giving the entire 31 question inventory, the team selected the five questions that best fit their needs. The student survey was only given to the identified gifted students within the classrooms. The results showed that: (a) 61% of the students reported being given the opportunity to select challenging books or assignments; (b) 79% reported being allowed to work at their own pace; (c) 89% experienced joy when working on challenging tasks; (d) 79% reported having classmates who understand how they learn; and (e) 75% reported being given interesting and challenging work.

The parent survey contained seven questions designed to elicit evidence of differentiation observed within the classroom by parents. The results showed that: (a) 46% of parents reported their child received a challenging curriculum; (b) 51% reported their child's strengths were addressed in the classroom; (c) 43% reported their child's interests were accommodated; and (d) 33% reported seeing evidence of differentiated work coming home. Comments left by the parents included several requests for more information about the cluster model and anecdotes or stories regarding individual student's experiences in the classroom.

Once all of the data were collected, the team identified several patterns and trends as directed in step eight of the DC-PET. First, the evaluation team noticed a discrepancy between the students' perceptions of the degree to which differentiation was being implemented and the parents' perceptions. Second, the team noted a general consensus among teachers that more time to plan and additional resources were needed. Third, several teachers' comments reflected the belief that the cluster model was not always being implemented with fidelity.

The team identified several strengths of the cluster model as directed by step nine of the DC-PET as well. First, the team listed that a majority of the gifted students served in cluster classrooms reported being challenged academically. Second, the team noted that a majority of teachers reported being well trained in differentiation and the cluster model. Third, the team listed the awareness of administrators to the importance of classroom composition in the success of the cluster model.

The recommendations created by the team in step 10 of the DC-PET were sent to the Director of Gifted Education. The recommendations aimed at administrators included: (a) using a checklist with indicators of differentiation during walkthroughs; (b) including differentiation in the professional learning communities used within the district; (c) encouraging the teachers under their charge to attend the cluster teacher meetings offered; and (d) consistently and carefully balancing the classroom composition in order to further narrow the range of instructional levels identified within the classroom.

The recommendations created by the team for gifted cluster teachers included: (a) encouraging each teacher to post information about the cluster model in classroom

newsletters and class websites; (b) exploring the resources available at the district gifted program office, as well as those offered online on the district website; and (c) providing additional time for cluster teachers to collaborate. Recommendations targeted towards parents included: (a) creating a district level quarterly parent newsletter with articles and updates; and (b) publicizing and encouraging parents to attend the gifted informational nights offered throughout the year.

After receiving the recommendations described above, the Director of Gifted Education converted them into SMART goals. Those goals state:

- The first issue of a quarterly parent newsletter will be created and distributed to parents within two weeks time.
- A new online cluster teacher platform for helping cluster teachers collaborate across the district will be put into place by the beginning of the next school year. Teachers using it will receive professional development hours useful for recertification.
- A frequently asked questions sheet and a sample beginning of the year welcome letter will be created and distributed to cluster teachers by the first week in June.
- Lunchtime mini-trainings for administrators on how to correctly construct a cluster-grouped classroom will be offered at the beginning of the next school year.
- A lead gifted cluster teacher will be chosen at each school site to facilitate improved communication by the beginning of the next school year.

The Director of Gifted Education shared the results of the evaluation with district level administrators and the parent council. The director also enthusiastically supported the suggestion by one evaluation team member that the evaluation team be converted into a more permanent advisory committee.

Evaluating the DC-PET. At the conclusion of the evaluation, each team member was asked to take a post-study survey and participate in a focus group. The response rate for the surveys and focus group are found in Table 23. The evaluation team's mean rating of evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.4 (0.89) on the pre-survey, but 2.3 (0.58) on the post-survey. The evaluation team's mean rating of expertise in using the Kaplan Depth and Complexity Model on a four-point scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.6 (0.55) on the pre-survey and 2.75 (0.50) on the post-survey.

The mean rating of the degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at All) to 4 (A Lot) can be found in Table 24. The mean rating for all 10 empowerment evaluation principles was 3.25 or greater, which indicates the participants believed the principles were evident frequently or a lot. The lowest rated principle was the inclusion principle at 3.25 (0.50) and the highest rated principles were the accountability principle and the improvement principle at 3.75 (0.50).

Table 23

Survey and Focus Group Response Rates for Case Study #1

	<i>N</i>
Pre-Survey Only	1
Post-Survey Only	0
Matched Pre- and Post-Surveys	4
Focus Group Participants	5
Total	5

Note. Out of 5 team members

Table 24

Empowerment Evaluation Principles Ratings for Case Study #1

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	4	3.50	0.58
Inclusion	4	3.25	0.50
Democratic Participation	4	3.50	0.58
Community Knowledge	4	3.50	0.58
Evidence-based Practices	4	3.50	0.58
Accountability	4	3.75	0.50
Improvement	4	3.75	0.50
Organizational Learning	4	3.50	0.58
Social Justice	4	3.50	0.58
Capacity Building	4	3.50	0.58

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

The team members voiced the opinion during the focus group that the DC-PET achieved a nice balance between being too structured and too open-ended. However, one individual wrote on the post-survey that they would have appreciated more structure and guidance. The skills reportedly learned by the team members during the process included: (a) how to properly phrase an evaluation question; (b) how to conduct a video

chat; and (c) how to take a larger idea and narrow it down to its essence. Regarding the formation of proper evaluation questions, L.S. said, the following:

I thought it was interesting trying to get to the best possible question we could to collect the information that we wanted. I really learned a lot from that. You know, from some of the questions, that I thought might be good ones, we dissected them and I realized that maybe they didn't quite answer what we were looking for. (L.S., focus group, April 22, 2015)

Individuals also stated that when comparing the DC-PET to other forms of evaluation, the DC-PET had more structure, operated more smoothly, and provided more support throughout the process. One participant wrote on her post-study survey that, "It was one of the better tools that I have used" (A.M., post-survey, April, 2015).

When asked if the Depth and Complexity Model icons served an important function or if they felt simply like an add-on, the team's consensus was that the icons created a framework for the evaluation and helped the process flow nicely. The team was also asked to list what they liked best and least about the tool. The best aspects of the tool listed by the team included: (a) the ability to generate actual data; (b) the opportunity to work as a team; (c) the step-by-step nature of the tool; and (d) a perception that the DC-PET was clear and concise. The least favorite aspect of the tool listed by two team members was the time intensive nature of the process. Another individual felt there was too much information included in the online application. Her suggestion was to "cut to the chase" (K.C., focus group, April 22, 2015). The mean rating for how likely individuals would be to use the DC-PET again in the future on a

scale from 1 (very unlikely) to 7 (very likely) was 5.75 (0.96). This indicated the team was between somewhat likely and likely to use the tool again.

Case study #2. The self-contained gifted evaluation team from District A is described in case study two. This evaluation team met separately from the team described in case study one and was led by a different individual. The district website describes the self-contained gifted program as being designed for highly and profoundly gifted students in grades 1-6 who are working two or more grade levels ahead of their peers. Qualifying students must (a) score 97% or greater on a state approved gifted test and (b) receive scores on the state achievement test meeting or exceeding grade level standards. Five elementary schools throughout the district offer the self-contained gifted program. All self-contained gifted classes are multi-age.

Participants. The self-contained gifted evaluation team consisted of five teachers who teach self-contained gifted classrooms. At least one representative from four of the five schools was included on the team. The leader of the team solicited a participant from the fifth school, but was unsuccessful. Teaching experience among the team members ranged from 4 to 35 years. During the 2014-2015 school year, two teachers taught grades 1-2, two teachers taught grades 3-4, and one teacher taught grades 5-6. All evaluation team members were White women. Four of the five team members reported having no prior experience with program evaluation. The fifth team member described examining a district program for strengths and weaknesses for a graduate level course.

Pre-study survey results. Similar to case study one, the self-contained gifted evaluation team members agreed unanimously that the purpose of any program evaluation is to determine the effectiveness of a program in order to make

improvements. The adjectives used to describe the gifted program were overwhelmingly positive. Examples included multi-tiered, differentiated, rigorous, varied, inviting, and strong. The one negative adjective phrase listed was *unaligned across grade levels and schools*.

The mean rating of pre-existing knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.2 (0.45). This rating placed the team in the novice category. The mean team rating of pre-existing knowledge of the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.0 (1.23). This placed the team in the "Know what it is, but have never used it" category. The large standard deviation can be explained by the fact four of the team members rated their knowledge of the Kaplan Depth and Complexity Model as a one or two, but one team member rated their knowledge as a four.

The evaluation team members listed two concerns regarding participating in the study. Three of the five team members noted the time it would take to complete the DC-PET. The second concern, mentioned by one team member, was a feeling of anxiety due to the newness of the experience.

Experience while using the DC-PET. The self-contained gifted evaluation team met three times in person and three times online. The team reported spending approximately 12 hours to complete all 11 steps. During this time, the team completed 12 weekly status checks in addition to sending several emails directly to me. Seven of the status checks contained no questions or concerns. The questions that were sent to me included (a) an inquiry regarding how to use the workbook properly, (b) a request for advice regarding survey creation, and (c) a desire for clarification regarding the target

date for completing the DC-PET. I received no requests for additional in person meetings after the initial orientation.

The team members reported relying most heavily on the paper workbook except for in the beginning and during the data collection and analysis phase. The evaluation team reported working well together and found the experience enjoyable. When thinking about her fellow evaluation team members, B.J. said:

I thought they were a dream...I wouldn't say that we agreed on everything to start with, but we talked about it and said, well, what is this? What did you mean? There was clarification; it wasn't like we were all on the same page all the time. (B.J., focus group, April 23, 2015)

Her colleagues agreed with this assessment.

Description of the evaluation conducted by the team. District A's self-contained gifted evaluation team chose paradox as the purpose of the evaluation. This meant the team would attempt to uncover any weaknesses or contradictory practices. After brainstorming a list of topics to explore, the team settled on the following three evaluation questions:

- What areas of our curriculum are and are not articulated and aligned?
- What assessments are used to place and monitor students' performance at "two grade levels ahead" as described on the district website?
- What about our program demonstrates or does not demonstrate defensible differentiation for gifted students?

In the end, only questions one and three were addressed.

The evaluation team created an administrator survey, a teacher survey, and a student survey to answer the questions listed above. The number of respondents from each stakeholder group included 3 administrators, 9 teachers, and 157 students. The teachers reported inconsistent access to materials and a wide variance in curriculum alignment among grades and schools within the self-contained gifted program. The administrators reported a willingness to support articulation and consistency in the program, but indicated in the comment section of the survey that they did not know how to do this. The questions for the student survey came from Gentry and Gable's (2000) *My Class Activities* instrument. The survey revealed (a) 75% of the students reportedly enjoyed school, (b) 70% felt challenged by the curriculum, and (c) 49% felt they were offered a measure of autonomy in the classroom.

Once the data were collected, the team created four SMART goals, which they shared with the Director of Gifted Education. The four goals were:

- All self-contained gifted teachers will meet with other self-contained gifted teachers teaching the same grades at least once per quarter to plan and articulate curriculum starting next year.
- A paid committee of representatives from each grade level will create a standard list of materials for self-contained gifted classrooms by July 1, 2015.
- All self-contained gifted classrooms will have all materials on the standard list of materials for their grades by August 1, 2015.
- Mentorship will be provided to all self-contained gifted teachers beginning next school year.

Upon receiving these recommendations, the Director of Gifted Education took immediate action and obtained approval from the Superintendent to offer six half-day release times for collaboration and planning. The major goal of these meetings will be to create a horizontal map of content covered at each grade level and school. The teachers will also create a list of must-do's and may-do's in order to encourage consistency from site to site. An inventory of classroom materials has already begun.

Evaluating the DC-PET. After completing the DC-PET process, each team member was asked to take a post-study survey and participate in a focus group. The response rates for the surveys and focus group are found in Table 25. The mean evaluation team rating of pre-existing knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.2 (0.45) on the pre-survey and 1.8 (0.45) on the post-survey. This placed the team between novice and proficient. The mean rating of pre-existing knowledge of the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.0 (1.23) on the pre-survey and 2.2 (1.10) on the post-survey. This kept the team in the "know what it is, but do not use it" category. The large standard deviation can be explained by the fact four team members consistently rated their knowledge of the Kaplan Depth and Complexity Model at one or two, but one team member gave herself a rating of four on the pre-study and post-study survey.

Table 25

Survey and Focus Group Response Rates for Case Study #2

	<i>N</i>
Pre-Survey Only	0
Post-Survey Only	0
Matched Pre- and Post-Surveys	5
Focus Group Participants	5
Total	5

Note. Out of five team members.

When asked to rate the degree to which the DC-PET aligned with the 10 empowerment evaluation principles, the mean rating for the improvement principle was the highest at 3.4 (0.89). The mean rating for the social justice principle was the lowest at 2.8 (1.10). The large standard deviation can be explained by three individuals rating social justice at two and two individuals rating it at 4. The complete list of mean ratings and standard deviations can be found in Table 26.

Table 26

Empowerment Evaluation Principles Ratings for Case Study #2

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	5	3.40	0.55
Inclusion	5	3.00	1.00
Democratic Participation	5	3.20	0.84
Community Knowledge	5	3.00	1.00
Evidence-based Practices	5	3.40	0.89
Accountability	5	2.80	0.84
Improvement	5	3.60	0.89
Organizational Learning	5	3.00	0.71
Social Justice	5	2.80	1.10
Capacity Building	5	3.20	0.84

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

The evaluation team reported learning the following new skills after using the DC-PET: (a) how to write quality evaluation questions; (b) how to use new types of data collection strategies and pre-designed instruments; and (c) the importance of always having strong goals along with an agenda for every group meeting. The evaluation team members expressed the belief that the DC-PET provided a balanced and clear structure that was paced, embodied choice, and had a nice flow. For example, C.J. stated:

It was organized, but it left a lot of room to make choices and decisions that would suit the needs of what we wanted to study and know the results for all the stakeholders in our study group. It also had many resources embedded in each section to aid in the group's decision. (C.J., post-survey, April 2015)

When comparing the DC-PET to other forms of evaluation used in the past, A.W. wrote, "I am a novice, but the tool was very easy to use and logical to follow" (A.W., post-survey, April 2015).

Only one of the five team members reported using the icons included in the tool to guide her thinking. She attributed her usage to the fact that she uses the Kaplan Model in her classroom with students on a consistent basis. The remaining team members agreed with this statement from L.B: "I felt that the explanations for the icons were helpful, but the icons themselves didn't really matter to me. After we got started, I don't think I really thought about them to be honest" (L.B., focus group, April 23, 2005).

Suggestions for improving the DC-PET included: (a) providing additional space on the workbook for participants to record answers; (b) discussing the value of the Kaplan Depth and Complexity Model to the evaluation more fully in the beginning; and (c) scripting the audio recordings in the app instead of including digital files. The group

reported working very well together and appreciated the leadership of the designated team leader.

The mean rating for how likely the team members would be to use the DC-PET again in the future on a scale from 1 (very unlikely) to 7 (very likely) was 5.2 (1.30). Three team members chose 4 – 5 and two chose 6-7. This placed the team in the “somewhat likely” category. Overall, the team expressed the belief that the process was validating, interesting, and valuable. As A.W. stated, “It was really nice to see that we kept finding they [the students] were happy, they are engaged, they are being challenged” (A.W., focus group, April 23, 2015). B.J. wrote, “It allowed me to collaborate with gifted professionals and find the commonalities we had in our thinking about the gifted program along with ideas to make the program even better” (B.J., post-survey, April 2015).

Case study #3. The third case study describes the content replacement gifted evaluation team from District A. According to the district website, the term content replacement is used to describe a daily pull-out program for students needing additional challenge in math and/or reading. The content replacement option is offered to grade 4-6 students at all 30 elementary schools within the district.

The participants. The evaluation team consisted of one administrator, three teachers, and one parent. Midway through the process, one of the teachers on the team had to be replaced by a different teacher due to a health issue. Teaching and administration experience ranged from 25 to 34 years. The teachers on the team were content replacement teachers responsible for teaching math and reading honors classes to grade 4-6 students in the program. All participants on the team were White women.

One team member reported having no prior experience with evaluating a program, while the other four listed (a) reviewing curriculum, (b) evaluating an afterschool grant program, (c) participating in a gifted program evaluation, and (d) coordinating a homeless family study for the state.

Pre-study survey results. An analysis of the pre-study survey taken by the five team members revealed an agreement that the purpose of evaluation is to identify strengths and challenges within a program in order to improve. The adjectives used to describe the gifted program in the district included 13 positive words (e.g., comprehensive, innovative, engaging, varied, and challenging) and two negative. The negative adjectives used to describe the program were *complicated* and *misunderstood*.

All but one evaluation team member had no concerns or anxieties about participating in the DC-PET study. The fifth team member worried that it would be difficult to critically reflect on oneself. The mean rating of self-reported knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.8 (0.45). This placed the team between novice and proficient. The mean rating of pre-existing Kaplan Depth and Complexity Model knowledge on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 1.6 (0.89). This placed the team between "I've never heard of it" and "I know what it is".

Experience while using the DC-PET. The team met six times in person over the course of six months, but collaborated online using Google Docs on a weekly basis. Overall, the group reported spending between 10 - 12 hours to complete all 11 steps. The team completed 12 weekly status checks and wrote "none" 11 times in response to *What questions do you have?* The question that was submitted asked for advice on how

to deal with data from 30 different schools. My response was to sample the group and randomize who participates.

The team reported working well together and did not have any conflicts. They did admit being sidetracked once or twice, but only for a short period of time. The team was split as to whether they relied more heavily on the paper workbook or the online application. The consensus was both were important and served their purpose during different stages of the evaluation. One team member did not use the workbook, but typed her responses to each item into a Word document. I received no requests for additional in-person meetings following the orientation.

Description of the evaluation conducted by the team. The content replacement evaluation team quickly settled on “contribution” as the purpose of the evaluation, indicating a desire to examine the effectiveness of the program. They reported reaching this decision in a matter of minutes. The process of choosing the right evaluation questions took approximately four hours, however. The three evaluation questions the team decided on were:

- What do content replacement students gain that they wouldn't have if not in a content replacement program?
- How effective is the current content replacement structure in addressing student needs?
- Are the affective needs of students in the content replacement classes being addressed?

The team described this as being an iterative process where they went back and forth between steps 3-5 often to ensure that each question was important and would lead to valuable findings.

In order to answer the chosen evaluation questions, seven different surveys were created and distributed. Separate surveys were sent to elementary students, middle school students, parents, gifted specialists, elementary teachers, middle school teachers, and administrators. The team decided not to send the surveys to the stakeholders at all 30 schools, but instead purposively selected a representative sample of 10. The evaluation team also carefully randomized which individuals at each school were asked to complete the survey. A dice was rolled and landed on three. As a result, every third person on the roster was asked to participate. The response rates for surveyed stakeholders by group were (a) 21% of elementary students, (b) 53% of middle school students, (c) 43% of parents, (d) 90% of gifted specialists, (e) 57% of elementary teachers, and (f) 83% of middle school teachers.

When analyzing the data, the team reported noticing three clear patterns. First, the administrators consistently pointed to the quality of the gifted specialists as a strength and scheduling issues as the biggest challenge. Second, the gifted specialists believed their ability to accelerate curriculum and provide academic rigor were strengths and expressed the desire for more time to plan. Third, the lack of a strong writing component within the reading curriculum was identified.

The teachers surveyed believed the strengths of the program were rigor and increased pacing, but felt that additional attention should be placed on the fundamentals. Teachers also expressed the belief that the program relied too much on acceleration and

not enough on enrichment. The elementary and middle school students believed the biggest strength of the program was the skills they learned (e.g., critical thinking, study skills, problem solving, collaboration). Students in 4th-7th grade did not identify any challenges with the current program. The parents that were surveyed overwhelmingly agreed that the gifted specialists were the biggest strength of the program due to the fact that the unique needs of all students were addressed. The largest challenge identified by the parents was a need for more information about the options available to identified gifted students in the district.

The evaluation team created two SMART goals after examining the list of strengths and weaknesses identified through the surveys. Those two goals were:

- Content replacement teachers will investigate ways to develop and support the written communication skills of the elementary gifted students at the September 2015 gifted specialist meeting.
- Content replacement teachers will meet quarterly with grade level teams in grades 4-6 to open a dialogue, explain, and share information regarding gifted instructional practices and content addressed in the gifted classrooms.

The findings from the evaluation and the SMART goals listed above were shared with all content replacement teachers throughout the district during a special meeting.

The results of the evaluation were also shared with the Director of Gifted Education. She pledged to act on the SMART goals and suggested the content replacement specialists also find ways to communicate better with the regular classroom teachers. In addition, trainers from the College of William and Mary were contacted and scheduled to provide additional training on the units of instruction District A purchased

in an effort to increase the written communication skills of the students.

Evaluating the DC-PET. The content replacement team was asked to take a post-survey and participate in a focus group like the other teams from District A. The response rates for the surveys and focus group can be found in Table 27. The mean rating of evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.8 (0.45) on the pre-survey, but 2.5 (0.58) on the post-survey. The mean rating of expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 1.6 (0.89) on the pre-survey and 3.25 (0.96) on the post-survey.

Table 27

Survey and Focus Group Response Rates for Case Study #3

	<i>N</i>
Pre-Survey Only	1
Post-Survey Only	0
Matched Pre- and Post-Surveys	4
Focus Group Participants	4
Total	5

Note. Out of 5 team members.

The degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at All) to 4 (A Lot) can be found in Table 28. The lowest rated principle was the social-justice principle at 3.0 (0.82). The three highest rated principles were the inclusion principle, the improvement principle, and the organizational learning principle at 3.75 (0.50).

Table 28

Empowerment Evaluation Principles Ratings for Case Study #3

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	4	3.25	0.50
Inclusion	4	3.75	0.50
Democratic Participation	4	3.50	0.58
Community Knowledge	4	3.00	0.00
Evidence-based Practices	4	3.50	0.58
Accountability	4	3.50	0.58
Improvement	4	3.75	0.50
Organizational Learning	4	3.75	0.50
Social Justice	4	3.00	0.82
Capacity Building	4	3.25	0.50

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

The team listed the strengths of the DC-PET as: (a) being comprehensive and clear; (b) providing step-by-step instructions; and (c) fostering a focus on goals and outcomes. One team member likened the tool to a “laser pointer” (T.T., focus group, April 28, 2015). The team believed that the DC-PET provided the correct amount of structure considering they self-identified as novice evaluators. D.A. said:

Coming from the perspective of someone who has never evaluated a program of this magnitude and been responsible for reporting the information back to people that (sic) are important, the structure was very important to me because I knew I didn’t leave anything out. (D.A., focus group, April 28, 2015)

Three of the team members identified the time commitment involved in using the DC-PET as the biggest drawback. However, R.C. stated, “When I first saw the process, I was like, this looks like a lot of work. But, as we got into it, I can honestly say, I enjoyed it” (R.C., focus group, April 28, 2015).

When asked to describe how the DC-PET was different from other forms of evaluation one team member wrote that it was “comprehensive and very clear” (K.B., post-survey, April 2015). A second individual wrote, “I thought it was definitely one of the more useful evaluation tools I have used” (T.T., post-survey, April 2015). A third evaluation member stated that when reflecting on previous evaluation methods “in the past, the questions and the focus were provided. Preparing your own required more work, however the feedback was more focused and valuable” (D.A., post-survey, April 2015).

The fourth team member who responded to the survey reported not having anything to compare the DC-PET with, but mentioned during the focus group that while using the DC-PET “we were pulling from all stakeholders. Sometimes we don’t” (R.C., focus group, April 28, 2015).

All four team members present for the focus group agreed that the Depth and Complexity Model icons served a valuable function in structuring the discussions. The leader of the group stated:

I constantly went back to the icons and asked, what is it we are doing right now? We are looking at this piece. Or, what are we doing right now? We are looking at this piece. That was helpful to me to maintain my focus because I felt like going into the meetings, I needed to be like, ‘guys, this is where we are next’. (K.B., focus group, April 28, 2015)

T.T. echoed that sentiment saying, “I equate them to like driving. You need stop signs or stop lights to tell you what you are supposed to be doing. That’s what those were” (T.T., focus group, April 28, 2015).

No recommendations for changing the tool were listed by the team members on the post-study survey. Similarly, when asked during the focus group what the team would do differently if they were to use the DC-PET again, the team agreed that it worked well as is. All members of the team felt the team leader did a superb job guiding the group. The team listed three new skills learned as a result of going through the process: (a) writing evaluation questions, (b) randomizing a sample, and (c) staying focused on a larger purpose without getting side tracked.

When reflecting after the completion of the evaluation, two team members commented on how they appreciated the ability to present actual data as justification for making changes instead of relying on a gut feeling. K.B. said, “It is so powerful for you to be able to stand up and say, ‘here is the data that supports the why’” (K.B., focus group, April 28, 2015). The mean rating for how likely the content replacement team would be to use the DC-PET again in the future on a scale from 1 (very unlikely) to 7 (very likely) was 5.25 (0.50). This indicated the team was between “somewhat likely” and “likely” to use the tool again.

District B Case Studies

Case studies four through six describe the results of three evaluation teams created to evaluate the gifted services offered in District B. The district is comprised of 39 schools and has a total student population of 36,400 students, 77% of whom are White. The second largest racial group within the district is Hispanic (14%). District B received an ‘A’ on the 2013/2014 state report card generated by the Arizona Department of Education. Only 17% of the students receive free or reduced lunches. The gifted

services offered throughout the district include content replacement for math and reading, cluster grouping, advanced courses, AP/IB, and self-contained gifted classes.

Unlike District A, the teachers and administrators participating on the evaluation teams did not receive any compensation for their time. The three teams met together as a group to receive directions regarding tasks to complete at the beginning of each meeting and then broke up into work groups to complete the tasks. Case study four focused on the self-contained program, whereas case studies five and six focused on the content replacement program and the cluster grouping program respectively. I attended four meetings to monitor and support the three teams at the request of the gifted coordinator. Representatives from all three evaluation teams participated in a combined focus group with me at the conclusion of the study.

Case study #4. According to District B's gifted services' scope and sequence, the self-contained gifted program "provides all day differentiated learning experiences addressing the individual needs of highly gifted students" (Scope & Sequence, 2014, p. 5). The district offers a self-contained gifted music academy for 1st - 6th grade gifted students and a self-contained gifted STEM academy for 1st – 8th grade gifted students.

Participants. The self-contained gifted evaluation team was comprised of two teachers, two parents, and two eighth grade students in the program. The first teacher had 15 years teaching experience and taught fourth grade in the STEM academy during the 2014/2015 school year. The second teacher had 29 years teaching experience and taught seventh grade in the STEM academy during the 2014/2015 school year. One teacher, one parent, and both students were men. All evaluation team members were White. Only one individual reported having any previous experience with program

evaluation. Her experience consisted of participating in self-evaluation due to the adoption of a teacher effectiveness initiative implemented at a previous school.

Pre-study survey results. An analysis of the pre-survey results revealed an agreement among the team members that the purpose of program evaluation is to improve the effectiveness of a program by examining strengths and weaknesses. All of the adjectives used to describe the gifted services offered in the district by the team members were positive. A few examples included diverse, comprehensive, awesome, challenging, evolving, and fun.

The mean rating of evaluation team members' self-reported knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.0 (0.0). This placed the team at the novice level. The mean rating for pre-existing knowledge of the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time") was also 1.0 (0.0). This placed the team squarely in the "I've never heard of it" category. Four of the individuals reported having no anxieties or concerns regarding participating in the study. One team member expressed a concern about the time it would take and the second individual felt a general sense of uneasiness often experienced by those experiencing something new for the first time.

Experience while using the DC-PET. The self-contained evaluation team met together as a group with the other two evaluation teams from District B nine times between the months of January and May. At each meeting, the gifted coordinator provided an agenda for the teams and supported the participants as they worked. She reported spending on average three hours in preparation for each one-hour meeting. The

group as a whole reported spending between 15 to 16 hours working to complete all 11 steps.

The evaluation team members indicated they worked well together, one individual saying, “Great collaboration. We do great together” (K.P., focus group, May 19, 2015). The self-contained evaluation team completed two status checks, both without any questions. The team reported relying more heavily on the paper workbook, but used the links within the app at the beginning and during data collection and analysis.

Description of the evaluation conducted by the team. The self-contained gifted program evaluation team decided the purpose of the evaluation would be to uncover weaknesses or contradictory practices within the program. The final three evaluation questions developed by the team were:

- How does the curriculum used in the self-contained gifted program translate into the high school years?
- Are students screened effectively to ensure that all students, especially underserved populations, are identified and served properly?
- Are we meeting the social and emotional needs of the students in the self-contained program?

Questions one and two were chosen from their list of unanswered questions. The third evaluation question was developed to address an equity and excellence issue. In the end, only questions one and three were addressed.

The team created three surveys to answer the evaluation questions above. The first survey was designed to measure the degree to which students’ social and emotional

needs were being met and was sent to current self-contained students. The second survey was designed to measure classroom quality based on Gentry and Gable's (2000) *My Class Activities* instrument and was sent to past participants in the program now in high school. The third survey was designed to measure the high school teachers' perceptions of the self-contained program after receiving past students in the program. The number of respondents from each stakeholder group included (a) 97 current students in the program, (b) 7 past students now in high school, and (c) 23 high school teachers. In addition, the team planned to interview the coordinator of the gifted and examine the gifted handbook to help answer question two. This did not actually take place, however.

The results of the survey given to current students showed: (a) 90% would recommend the program to a friend; (b) 86% found the classwork challenging; (c) 70% believed role models who matched their abilities and strengths were introduced; and (d) 60% felt the program encouraged creativity. Interestingly, 73% of students felt comfortable around their self-contained gifted peers as compared to only 43% around general education peers.

The teacher survey revealed five major findings as reported by the high school teachers who receive self-contained gifted students when they age out of the self-contained gifted program. Those findings were: (a) gifted students have limited access to technology during their high school years; (b) students need additional social and emotional support during high school; (c) the high school teachers are unaware of the degree to which students in the program were accelerated up to the eighth grade; (d) high school teachers do not feel comfortable providing high levels of differentiation

themselves; and (e) few high school teacher are consistently using gifted strategies in order to meet their students' needs.

In order to mediate the weaknesses described above, the team developed the following four SMART goals:

- High school teachers will promote student use of personal technology including phones, tablets, and laptops starting with the 2015-2016 school year in order to increase access to technology.
- The district will purchase a curriculum to meet the social and emotional needs of 1st-8th grade self-contained students by August 2015.
- Discussions will take place beginning 2015-2016 between the self-contained gifted teachers and the high school counselors to bridge the gap between 8th grade self-contained students and their future high school teachers.
- The district will offer gifted classes and professional development opportunities for high school teachers during the 2015-2016 school year. The goal will be to have at least one teacher representative per department at each high school attend the 'Gifted 101' class with the expectation that they share the information with the rest of their department.

The coordinator of the gifted program was supportive of the team's goals and will present the findings in a formal evaluation report to all relevant district level administrators and board members.

Evaluating the DC-PET. The response rates for the surveys and focus group can be found in Table 29. The evaluation team's mean rating of evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.0 (0.0) on the pre-survey, but 2.0 (0.0) on the

post-survey. This indicated a jump from novice to proficient. The mean rating for pre-existing expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 1.0 (0.0) on the pre-survey and 2.67 (0.58) on the post-survey.

The mean ratings for the degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at all) to 4 (A Lot) can be found in Table 30. The democratic participation principle, the community knowledge principle, and the improvement principle were rated the highest at 3.33 (0.58). The accountability principle and the social-justice principle were rated the lowest at 2.67 (1.16). The high standard deviation can be explained by two team members rating social justice at two and one team member rating it at four.

Table 29

Survey and Focus Group Response Rates for Case Study #4

	<i>N</i>
Pre-Survey Only	3
Post-Survey Only	0
Matched Pre- and Post-Surveys	3
Focus Group Participants	3
Total	6

Note. Out of 6 team members.

Table 30

Empowerment Evaluation Principles Ratings for Case Study #4

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	3	3.00	1.00
Inclusion	3	3.00	1.00
Democratic Participation	3	3.33	0.58
Community Knowledge	3	3.33	0.58
Evidence-based Practices	3	3.67	0.58
Accountability	3	2.67	1.16
Improvement	3	3.33	0.58
Organizational Learning	3	3.00	1.00
Social Justice	3	2.67	1.16
Capacity Building	3	3.00	1.00

Note. Response scale ranged from 1 (Not at All) to 4 (A Lot)

When asked to describe what new skills were acquired while using the DC-PET the teams reported learning how to create SMART goals and the importance of wording questions in a survey properly. H.S. stated, “The verbiage on one wasn’t exactly what we wanted, so what we got was a different answer. We had to re-do it” (H.S., focus group, May 19, 2015).

Three of the team members were present for the combined focus group conducted for District B. One of the three team members reported not using the Kaplan icons embedded in the tool. He stated:

I didn’t spend a lot of time on it. I would look at the icon, and I tried to figure out, how does this represent what I am looking it? It didn’t make sense to me. I went, that’s not what I would picture as that and that didn’t mean anything to me. (H.S., focus group, May 19, 2015)

He explained that the concepts the icons represented made sense to be used with an evaluation and were valuable, but the pictures themselves did not seem to represent the concept.

The evaluation team believed the DC-PET had the right amount of structure and enjoyed the flexibility it provided. K.P. stated, “The structure was perfect” (K.P., focus group, May 19, 2015). The parent on the team went on to say:

They actually let us choose where it is we wanted to go. What are those strengths and those weaknesses? Identify them. How are you going to prove that you’ve met those particular things? That shows they have value or trust in us. (J.S., focus group, May 19, 2015)

The team also spoke about the usefulness of the DC-PET to identify more complex issues within the program.

The members of the self-contained gifted evaluation team recommended that a thorough overview of the DC-PET process be included in the online app instead of simply using the PowerPoint presentation I used at the first meeting. They believed the tool was confusing at first, but became clearer as the process continued. K.P. said, “There’s a point for me when the light bulb went off and I was like, ‘Oh my Gosh’. This was really great and we’re going to have really good results” (K.P., focus group, May 19, 2015).

The team lost one parent and one student from the team by the end of the evaluation. When I asked the team to reflect on why this happened, one theory was that “Life got in the way” (J.S., focus group, May 19, 2015). The time commitment seemed too much for some individuals. The second student on the team did not complete the

post-survey either, but attended most meetings with his mother. The only active role he took was in helping to turn the paper survey chosen by the other members of the team into a Google form. When thinking about this, one member stated, “I would like to know from the beginning, how could we get the students’ perspective without necessarily making them sit through these meeting?” (H.S., focus group, May 19, 2015). The group valued the students’ perspective, but did not necessarily think having students be formal members of the evaluation team was appropriate.

If the team were to use the DC-PET again, they would spread out the implementation over a longer period of time instead of just six months. They also discussed the best time of year to start. The final solution was to wait until the beginning of the year chaos was over, and finish before the end of the year chaos began. The team thought it would be a good idea to set meeting dates for the entire year at the beginning instead of proceeding week to week. The mean rating for how likely they would be to use the DC-PET again in the future on a scale from 1 (very unlikely) to 7 (very likely) was 4.67 (0.58). This indicated the team was between “undecided” and “somewhat likely” to use the tool at some point again in the future.

Case study #5. Case study five focused on the content replacement program evaluation team from District B. The content replacement program consists of providing accelerated curriculum in English language arts and math to qualifying students in grades 3-6 outside of the regular classroom (Gifted Services Handbook, 2014, p. 10). The students attend the regular classroom for all other subject areas.

Participants. The team consisted of three teachers, one parent, and one administrator. Teaching and administration experience ranged from 7 - 23 years. All

content replacement teachers teach math and language arts to the students identified for the program at their school. The participants on the team were White women except for the parent. The father had a student in both the self-contained program and the content replacement program. One member reported having no previous experience with program evaluation. The remaining team members listed (a) being a part of an English language learner's task force, (b) serving on an accreditation team, (c) evaluating personnel, and (d) reviewing a program with no formal training.

Pre-study survey results. The team agreed that the purpose of evaluation is continuous improvement in order to maximize effectiveness based on strengths and weaknesses. Nine of the adjectives used to describe the program were positive (e.g., growing, rich, rigorous, fluid, flexible). The two adjectives of a negative nature included in the list were *disjointed* and *inconsistent*, both of which were listed by the same person.

The mean rating of self-reported knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.8 (0.75). This placed the team between the novice and proficient level. The mean rating of pre-existing knowledge of the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 1.3 (0.52). This placed the team in the "I've never heard of it" category.

One team member reported having no anxieties or concerns regarding participating in the study. Three team members expressed concern over the time it would take to complete the process. The last individual had a fear that she might not be able to contribute much due to a lack of experience with evaluation.

Experience while using the DC-PET. The content replacement evaluation team met together as a group with the other two evaluation teams from District B. The team reported spending 15 - 16 hours working on the process. When asked to describe how well they worked together, C.G. stated, “Well, a lot of our group dropped out. Apparently we were not easy going...I don’t know” (C.G., focus group, May 19, 2015). The reasons for why several members decided not to continue on the team are explored in the *Evaluating the DC-PET* section to follow.

The team members completed six status checks, only asking procedural questions such as, “Do I have to complete something online for step one parts A and B or just in the book?” The two members of this team present for the combined District B focus group reported using the electronic app more frequently in the beginning, but the paper workbook more often after that.

Description of the evaluation conducted by the team. The content replacement team decided the purpose of their evaluation would be to determine the effectiveness of the program and uncover weaknesses or contradictory practices. The final three evaluation questions developed by the team were:

- Do our current programs address the social emotional needs of gifted learners? If so, how?
- Is the gifted education English language arts curriculum used in grade 3-5 consistent, rigorous, and aligned to the new state standards?
- How do we meet the academic and social/emotional needs of twice-exceptional learners?

Questions one and two were chosen from the list of unanswered questions. Question three came from the list of equity and excellence issue included in the tool. In the end, questions one and two received a majority of the focus.

The team chose to develop a survey and conduct a document review in order to answer their questions. The survey created by the team focused on social and emotional needs of students and was sent to the parents of all content replacement students. The survey had 10 questions and a response rate of 97 individuals. Findings showed: (a) 46% believed the program adequately or thoroughly provided students an opportunity to explore social and emotional concerns; (b) 80% believed the program encouraged enthusiasm for learning; and (c) 70% believed the program allowed students to explore their interests and strengths.

The document review undertaken by the team involved comparing the district created gifted curriculum guide with (a) the Common Core State Standards and (b) the College of William and Mary curriculum units adopted by the district. This correlation uncovered numerous skills not being addressed by the curriculum that were included in the standards and the district curriculum guide. One team member remarked that she was not surprised by this finding and reflected on how the content replacement teachers often filled in the gaps on their own with no consistency from site to site (C.G., focus group, May 19, 2015).

The team created two SMART goals in an attempt to address these findings. The goals were:

- District B will purchase appropriate ELA curriculum resources to be used with consistency and aligned to the new state standards in order to address the needs of the gifted students in grades 3-5 by August 2015.
- All schools will provide quarterly social/emotional support opportunities for gifted students with the gifted specialist and school counselor using an adopted program during the school day focusing on self understanding of strengths and weaknesses characteristic of a gifted student using NAGC standards by next school year.

The gifted coordinator was supportive of the team's goals and planned to share the results with district-level administrators.

Evaluating the DC-PET. The response rates for surveys and focus group can be found in Table 31. The mean rating for existing evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.8 (0.75) on the pre-survey, but 3.0 (0.0) on the post-survey, indicating a jump from novice to advanced. The mean rating of knowledge regarding the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 1.3 (0.52) on the pre-survey, but 3.5 (0.71) on the post-survey. This indicated a move from "I've never heard of it" to "I've used it a few times in the past".

The mean ratings for the degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at All) to 4 (A Lot) are found in Table 32. All principles were rated at the highest level of 4.0 (0.0) except for the inclusion principle, which was rated at 3.5 (0.71).

Table 31

Survey and Focus Group Response Rates for Case Study #5

	<i>N</i>
Pre-Survey Only	3
Post-Survey Only	0
Matched Pre- and Post-Surveys	2
Focus Group Participants	2
Total	5

Note. Out of 5 team members.

Table 32

Empowerment Evaluation Principles Ratings for Case Study #5

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	2	4.00	0.00
Inclusion	2	3.50	0.71
Democratic Participation	2	4.00	0.00
Community Knowledge	2	4.00	0.00
Evidence-based Practices	2	4.00	0.00
Accountability	2	4.00	0.00
Improvement	2	4.00	0.00
Organizational Learning	2	4.00	0.00
Social Justice	2	4.00	0.00
Capacity Building	2	4.00	0.00

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

The post-study surveys for the content replacement team indicated a belief that the process was a valuable experience. When describing the process of reflecting, questioning, collecting, and analyzing data, one member wrote, “This is not a process I would have been able to do before the study” (K.L., post-survey, May 2015). The team also appreciated being able to work with peers and collaborate. Individuals stated that

when comparing the DC-PET to other forms of evaluation, this actually lead to goals with action steps. K.L. stated, “This process was much more intentional and structured specifically for gifted programming...The other one I did (sic) was dropped on me and was pointless and had meaningless results” (K.L., post-survey, May 2015). A second member wrote, “The DC-PET tool was very detailed and easy to follow, getting results for our program and information that will improve our practices” (C.G., post-survey, May 2015).

Both team members present for the combined focus group in District B thought that the Kaplan Model icons served an important function. C.G., said, “It gave credibility to the process. It’s linked to something that’s credible” (C.G., focus group, May 19, 2015). K.L. responded by saying, “For me, it peaked my interest about Kaplan...It is something I will explore because of this experience” (K.L., focus group, May 19, 2015).

The team lost one parent and two teachers by the end of the evaluation. When I asked the team to reflect on why this occurred, the number one theory was that some people did not understand the intent of the tool. One member of the team stated:

I went into it thinking it was an assessment piece that was already created for us to implement and we were going to learn how to implement it. And, when I came to the first meeting and realized, oh, we’re actually going to create this, not just implement it. Maybe explaining this better initially would have kept people from dropping. (C.G., focus group, May 19, 2015)

The administrator on the team also said, “I heard somebody say that they thought it was going to be an opportunity for them to express their concerns... and they didn’t want to

do the work” (K.L., focus group, May 19, 2015). The mean rating for how likely the team would be to use the DC-PET again in the future on scale from 1 (very unlikely) to 7 (very likely) was 6.5 (0.7). This indicated the team was between “likely” and “very likely” to use the tool again in the future.

Case study #6. Case study six focused on the cluster grouping evaluation team from District B. The gifted services scope and sequence developed by the coordinator of the gifted program within District B describes cluster grouping as, “a method of service in which four to eight gifted students are placed together in a classroom with other non-gifted students and a teacher trained in differentiating for gifted learners” (Scope & Sequence, 2014, p.4). The members of this team met simultaneously with the other two evaluation teams within District B, but completed tasks independently at a different table.

Participants. The cluster grouping evaluation team consisted of three teachers and one administrator. Teaching and/or administrative experience ranged from 14 to 23 years. One teacher taught first grade as a cluster teacher and two teachers taught grades 4 through 6 math and reading enrichment for clustered students during the 2014/2015 school year. All but one individual were woman. All participants were White. Prior experiences with program evaluation listed by the participants included (a) leading a school improvement team, (b) evaluating teachers at a preschool, and (c) helping with a gifted program evaluation 10 years in the past.

Pre-study survey results. The four members of the team expressed similar beliefs that the purpose of evaluation is to make improvements and gain knowledge. One individual went on to express the notion that evaluations should represent the views of

all stakeholders (S.B., pre-survey, January 2015). All adjectives except for one used to describe the gifted program in District B were positive. Examples included growing, inclusive, improving, diverse, and research-based. The one negative adjective used to describe the program was *limited*.

The mean rating of each team members' self-reported knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.67 (0.58). This placed the team between the novice and proficient level. The mean rating for pre-existing knowledge of the Kaplan Depth and Complexity Model on scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.0 (1.0). This placed the team in the "Know what it is but haven't used it" category.

One individual reported having no anxieties or concerns about participating in the study. The anxieties reported by the remaining team members included (a) time constraints, (b) a possible lack of support, and (c) worrying that the tool might not do what it was designed to do, therefore wasting their time.

Experience while using the DC-PET. The cluster grouping evaluation team met together as a group with the other two evaluation teams from District B. The team reported spending 15 – 16 hours working on the DC-PET. No conflicts were reported between evaluation team members. In fact, M.C. said, "I thought we did great. I enjoyed working with everyone" (M.C., focus group, May 19, 2015). One individual on the team described using the online application 70% of the time and the workbook 30% of the time. The remaining members of the team present during the focus group favored the workbook over the electronic app.

The cluster grouping evaluation team completed four status checks. Only two questions were submitted. The first question was a procedural question regarding how to use the DC-PET workbook. The second question did not pertain to the study, but inquired about the best way to meet the needs of each individual gifted learner in a classroom.

Description of the evaluation conducted by the team. The cluster grouping program evaluation team members chose to measure the effectiveness of the program. The final three evaluation questions developed by the team were:

- How do we know the cluster model is effective at meeting the academic needs of our gifted students?
- Are we implementing the essential elements of an effective cluster model?
- How do we know the cluster model is effective at meeting the social emotional needs of our gifted students?

Questions one and two came from the self-generated list of unanswered questions.

Question three was selected from the list of equity and excellence issues included in the tool.

In order to answer the evaluation questions, three surveys were developed and administered. Seventeen students responded to a student survey, 20 teachers responded to a teacher survey, and 36 parents responded to a parent survey. The student survey consisted of selected questions taken from Gentry and Gable's (2000) *My Class Activities* instrument. The results indicated: (a) 70% of students have friends at school, (b) 68% believe the teacher cares about their ideas or concerns, (c) 56% believe the

teacher provides choices, and (d) 37% believe classroom work is challenging or interesting.

A majority of the teacher survey questions were created by rewording the questions on the student survey to focus more on the teacher. For example, a question on the student survey said, *My teacher checks to see what I already know before teaching me something new*. The corresponding teacher survey question was, *I pre-assess my students to drive instruction*. The teacher survey revealed: (a) 90% of teachers believe students have friends at school; (b) 95% take time to listen to students' concerns; (c) 60% address different learning preferences; (d) 90% believe they provide their students with the appropriate challenge level; (e) 75% of teachers feel no one else on campus meets the social and emotional needs of gifted students except for them; and (f) 75% of gifted cluster teachers believe they need more time to collaborate with other teachers during the school day.

The parent survey was constructed in a similar manner to the other two surveys. For example, the same pre-assessment question described in the paragraph above was rewritten for the parent survey to read, *My child's teacher uses assessments to plan the lessons*. The survey revealed: (a) 72% believe students have friends at school, (b) 61% believe their child can share concerns with their teacher, (c) 44% believe the teacher offers choices, and (d) 66% believe their child is challenged by the curriculum at the correct level. The most surprising finding to the team was that 60% of parents reported not knowing about the gifted services provided throughout the district.

Three SMART goals were created by the team to address the findings described above. The goals were:

- All gifted teachers will be given at least three opportunities during contract hours to collaborate with teachers of similar grade levels or subject areas on meeting the needs of gifted students beginning the next school year.
- All schools will provide quarterly social/emotional support opportunities for gifted students with gifted specialist/counselor using an adopted program during the school day focusing on self understanding and gifted characteristics using NAGC standards beginning next year.
- Gifted cluster teachers will communicate with parents at the beginning of the school year and quarterly thereafter regarding gifted practices and differentiation in their classroom with planning time from administration to complete the preparation of materials during contract hours.

The SMART goals only addressed two of the three original evaluation questions the team developed. The gifted coordinator is supportive of their efforts and plans to address their concerns in the near future.

Evaluating the DC-PET. The response rates for the surveys and focus group are found in Table 33. The mean rating of the evaluation team's evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.67 (0.58) on the pre-survey, but 2.33 (0.58) on the post-survey, indicating a jump from novice to proficient. The mean rating for expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.0 (1.0) on the pre-survey, but 3.5 (0.71) on the post-survey.

Table 33

Survey and Focus Group Response Rates for Case Study #6

	<i>N</i>
Pre-Survey Only	3
Post-Survey Only	0
Matched Pre- and Post-Surveys	3
Focus Group Participants	3
Total	3

Note. Out of 4 members.

The mean ratings for the degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at All) to 4 (A Lot) are found in Table 34. The two principles rated the highest were the inclusion principle and the democratic participation principle at 3.67 (0.58). The lowest rated principle was the social-justice principle, which was rated at 2.67 (1.16). The large standard deviation can be explained by two individuals rating social justice at 2 and one individual rating it at 4.

Table 34

Empowerment Evaluation Principles Ratings for Case Study #6

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	3	3.00	0.00
Inclusion	3	3.67	0.58
Democratic Participation	3	3.67	0.58
Community Knowledge	3	3.33	0.58
Evidence-based Practices	3	3.00	1.00
Accountability	3	3.33	1.16
Improvement	3	3.33	0.58
Organizational Learning	3	3.33	0.58
Social Justice	3	2.67	1.16
Capacity Building	3	3.00	1.00

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

The post study surveys for the cluster grouping evaluation team indicated a belief that the DC-PET was an organized, in-depth process that broadened the participants' perspectives. One individual wrote that the process, "seemed overwhelming at first, but was simple when we broke it down step-by-step" (M.C., post-survey, May 2015). Individuals on the team stated that when comparing the DC-PET to other forms of evaluation, the DC-PET, "had more of a whole or big-picture theory" (S.B., post-survey, May 2015) and "used similar processes, but broadened our perspective" (M.C., post-survey, May 2015).

The team had mixed opinions on the usefulness of the icons. As an example, one individual said, "I think it's half and half. I liked some of the icons because I'm a visual learner, so for some of them it helped me remember" (S.B., focus group, May 19, 2015). The number one skill the team believed they learned from the process was how to write SMART goals. Regarding this, A.L. explained that in a previous evaluation attempt, "we just collected information and then we were done. Nothing changed. Like, I don't know what the outcome was" (A.L., focus group, May 19, 2015). The consensus was that developing SMART goals would ensure real change actually took place.

The mean rating for how likely the team members would be to use the DC-PET again in the future on a scale from 1 (very unlikely) to 7 (very likely) was 5.0 (0.0). This indicated the team was "somewhat likely" to use the tool at some point again in the future.

District C Case Study

The seventh case study originated in District C. This district has over 27,000 students, 37 schools, and received an A on the 2013/2014 state report card generated by the Arizona Department of Education. Almost three quarters of the students were White. The second largest racial group was Hispanic (17%). Only 22% of the student body received free or reduced lunch. The gifted program in this district is large and offers a variety of services including grade acceleration, cluster grouping, self-contained gifted classes, and AP / IB. Despite the fact that District C is highly performing, financial stresses abound and charter schools have been attracting students out of the system.

Case study #7. The evaluation team created in District C consisted of individuals from a high performing school within the district. In fact, according to the school's website they are ranked in the top 15 American public schools in the nation. The method of service currently used to meet the needs of gifted students enrolled in the school is a daily pull-out model that focuses on math and integrating the curriculum.

Participants. The evaluation team consisted of 11 stakeholders. Four of the team members were teachers, four were parents (two with gifted children in the program and two without), two were administrators, and one was a community member. All participants were women except for one of the two administrators. Teaching and administration experience ranged from 1 to 33 years. One of the teachers was the gifted pull-out teacher for grades 3 through 5 and the others were regular education teachers that send students to her each day. All participants were White. Six of the participants reported having no previous experience evaluating a program. The remaining participants listed: (a) participating in a gap analysis; (b) facilitating an evaluation of

school-wide behavior program; and (c) piloting and gathering data on a new method of classroom instruction.

Pre-study survey results. An analysis of the pre-study survey taken by nine of the 11 individual team members revealed a general consensus that the purpose of evaluation is to examine the effectiveness of a program and to improve over time. The team also agreed that the gifted program in District C was comprehensive. That is where the agreement stopped. Thirteen of the adjectives used to describe the program were positive (e.g., inclusive, varied, rigorous, creative, student-centered) and 11 of the adjectives were negative (e.g., exclusive, secretive, underfunded, inequitable, dull). The mean self-reported rating for the amount of existing program evaluation knowledge possessed by the group on a scale from 1 (Novice) to 4 (Expert) was 1.33 (0.5). This placed the team between the novice and proficient categories. The mean rating for pre-existing Kaplan Depth and Complexity Model knowledge on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.4 (1.13). This placed the team between "I know what it is" and "I've used it once or twice". The high standard deviation can be explained by the fact five team members rated their knowledge of the model at a one or two. Four team members rated their knowledge at a three or four.

Two of the team members reported having no anxieties or concerns about participating in the study. A second pair of individuals felt that time was the biggest issue. A third pair of team members were afraid the data collected would not be used to improve the program despite everyone's best efforts. One individual worried that resentments might develop if the data were used to create recommendations. The last

participant was new to the school and feared that she did not have enough knowledge of the district and the gifted program to be of much help.

Experience while using the DC-PET. The team members met in person 12 times over the course of five months. I was present at four of the evaluation team meetings per the request of the team lead. All members agreed that the electronic app was most useful at home and the paper workbook was used more often during in-person meetings. The team reported spending about 14 hours on the process, but they did not complete all 11 steps. The team was not able to come to agreement on what the three evaluation questions should be and cited irreconcilable differences between the teachers on the team and the parents as the reason. This unfortunate development will be explored in a later section of this case study.

The team completed four online status checks during the process. The first two status checks reported the team was working well and no questions were asked. The last two status checks reported major problems. Issues mentioned were (a) disorganized discussions, (b) one or two people dominating the conversation, and (c) fear that everyone was operating with their own personal agendas. In response to these concerns, I replied personally to the individuals submitting the status checks and then contacted the Gifted Coordinator to see what could be done to alleviate tensions.

Description of the evaluation conducted by the team. As mentioned above, the team from District C did not complete an evaluation due to issues that arose between the parents and the school staff. Midway through the evaluation an email was sent from the parents to the teachers and administrators explaining that the parents had decided not to participate any longer. The email attributed the decision to leave the process to a feeling

by the parents that the educators had more experience and therefore best served the evaluation. The parents wrote that they believed their contribution would be minimal, at best. The parents then provided four suggestions for changing the current program and wished the team well. Two additional meetings with only teachers and administrators present took place after the parents' exit before the decision was made to stop the evaluation completely.

Altogether, the team spent five of the 12 meetings attempting to develop evaluation questions. I came to give a mini-workshop on best practices for writing evaluation questions in an effort to help and offered to share additional example questions including those used in the Arkansas Evaluation Initiative templates (Robinson, Cotabish, Wood, & Biggers, 2009). It was obvious to me at that meeting that a stalemate had developed and the morale of the group was low.

According to an extensive discussion during the focus group, the teachers and administrators believed that the parents had an agenda and were not open-minded. For example, J.S. stated:

I just felt that some of the parents had an agenda and they wanted to address their agenda and what they wanted, not necessarily what the research showed or what other community or staff members wanted. Some, not all. (J.S., focus group, April 27, 2015)

A second teacher confirmed this by saying, "I kind of felt like some of them weren't open to evaluating what's currently going on. Because, they'd already decided" (P.H., focus group, April 27, 2015).

The teachers also discussed the fact that, in their minds, most of the possible evaluation questions being generated by the parents were leading and showed a lack of knowledge about the gifted program as a whole. T.H. said, “The meetings did feel very different depending on who showed-up” (T.H., focus group, April 27, 2015). J.S. added to the conversation by saying, “One particular member was very vocal and wanted what we had before and then seemed to get other people to jump on the band wagon” (J.S., focus group, April 27, 2015).

Several of the team members also thought that the parents wanted immediate change and would not be patient for results. As an example, C.Y. said:

I think we tried to set that up in the beginning to say we’re here just to figure out what’s working well and what can we improve upon and create a three to five year plan. It’s not going to be an immediate change for the fall and I think that people got excited about the idea for new opportunities and expected things next fall. And then, when they started thinking about how the rubber would hit the road, that’s when I think we encountered some barriers. (C.Y., focus group, April 27, 2015)

P.H. confirmed this by saying, “I think you’re right. I think a lot of the parents were expecting more of a fast change” (P.H., case study, April 27, 2015).

The one parent who responded to the post-survey felt quite differently from the teachers. She said, “In my approximate six weeks of participation, much of the time has been spent off-topic...No one in the group has a true desire to follow the steps” (M.C., post-survey, April 2015). She went on to write, “When multiple members of the group state that parents are not educators and that parent expectations of the program are not

important, it is hard for parents to feel included or part of a democratic process” (M.C., post-survey, April 2015).

Communication was definitely a barrier. One comment in particular from the parent mentioned above stood out. She wrote:

I am struggling to understand how to reconcile the Gifted Coordinator’s idealized version of gifted education with the real world classroom experience. As a parent, how do I respond when she presents something that I know isn’t done in our classrooms? (M.C., post-survey, April 2015)

At the same time, the gifted teacher at the school began to feel like the parents were evaluating her and not the program. She wrote, “After three and a half months of nit picking, you start taking it personally” (J.S., focus group, April 27, 2015).

Another cause attributed to the failure of the evaluation was a district culture that encourages looking outward to find new ideas, not reflecting and introspectively looking inward. L.B. stated:

There was a dynamic here because so much of what we do, we’re so used to kinda (sic) looking outside and saying, “What are other schools doing?” and “How can we do something better or similar or different?” or “How do we stand out?”... This is the exact opposite of that. This is looking internally and looking at a program that already exists and not looking outside. (L.B., focus group, April 27, 2015)

Despite the fact that a successful evaluation did not take place, one of the administrators stated, “While we didn’t go through the steps, I think we also got a lot of answers,

because there were a lot of ideas and issues that were mulled over” (W.D., focus group, April 27, 2015).

Evaluating the DC-PET. The response rates for the surveys and focus group are shown in Table 35. The missing data shown in the table was due to the exit of the parents from the evaluation process. The mean rating for the evaluation team’s knowledge of program evaluation on a scale from 1 (Novice) to 4 (Expert) was 1.33 (0.5) on the pre-survey, but a 2.0 (0.58) on the post-survey. The mean rating for expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I’ve never heard of it) to 4 (I use it all the time) was 2.4 (1.13) on the pre-survey, but 3.14 (0.69) on the post-survey.

Table 35

Survey and Focus Group Response Rates for Case Study #7

	<i>N</i>
Pre-Survey Only	3
Post-Survey Only	1
Matched Pre- and Post-Surveys	6
Focus Group Participants	7
Total	10

Note. Out of 11 members.

The mean ratings for the degree to which the team felt the DC-PET aligned with the 10 empowerment evaluation principles are shown in Table 36. It is important to note that when rating the principles, one respondent wrote:

I am not confident that the answers I provided related to the specific principles is an accurate reflection of the power of the tool. I say this because we did not

complete the process so the tool was not used as intended. I do see how the tool would allow for the principles to be applied. (K.D., post-survey, April 2015)

The highest rated empowerment evaluation principles were the inclusion principle and the evidence-based practices principle, both at 3.14. The lowest rated was the organizational learning principle at 2.0 (0.82).

Table 36

Empowerment Evaluation Principles Ratings for Case Study #7

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	7	3.00	0.82
Inclusion	7	3.14	1.07
Democratic Participation	7	3.00	1.00
Community Knowledge	7	2.71	0.76
Evidence-based Practices	7	3.14	0.69
Accountability	7	2.42	0.79
Improvement	7	2.57	0.98
Organizational Learning	7	2.00	0.82
Social Justice	7	2.42	0.98
Capacity Building	7	2.42	0.98

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

The teachers on the team liked that the DC-PET provided a framework for discussion in the pursuit of improving the program. P.H. said, “The instrument forced me to be open-minded” (P.H., post-survey, April 2015). Three individuals stated that when comparing the DC-PET to other forms of evaluation, the DC-PET was more thorough, formal, and resembled an independent study project.

A suggestion offered for improving the DC-PET made by the one parent who completed the post-survey was to outline the goals and objectives more clearly so that everyone had similar expectations. This parent went on to say the process resulted in “spinning of wheels and very little forward progress” (M.C., post-survey, April 2015). The teachers and administrators suggested adding additional structure and support built into the process for those who are struggling or become stuck along the way. One teacher said the process was too open ended, and another believed the DC-PET assumed a level of familiarity with the process of evaluation that she did not have.

The positive outcomes listed by the teachers and administrators due to participating in the study included learning how to (a) form evaluation questions, (b) conduct a focus group, and (c) deal with conflict. In addition, a consensus developed among the teachers and administrators that they needed to improve how they communicate with parents. Suggestions generated by the teachers and administrators for what the team could have done differently if they were to use the DC-PET again in the future included: (a) setting norms and hard deadlines for each meeting; (b) not starting in the middle of the school year; (c) creating a smaller evaluation team; and (d) designating one person to be the leader/facilitator who would then summarize and chart the progress for the team before and in-between each meeting.

The mean rating for how likely the team would be to use the DC-PET again in the future on a scale from 1 (very unlikely) to 7 (very likely) was 4.14 (1.77). This indicated the team was between “undecided” and “somewhat likely” to use the tool at some point in the future. The large standard deviation can be explained by three

individuals responding with 2 or 3, two individuals responding with 4, and two individuals with 6 or 7.

District D Case Study

The eighth case study took place in District D. Unlike districts A, B, and C, this district had: (a) 64% of the students eligible for free or reduced lunch; (b) limited gifted services in the form of cluster grouping and a once a month after school enrichment activity; and (c) a mostly non-White student body (60% Hispanic). The district consists of nine schools and received a C on the 2013/2014 state report card generated by the Arizona Department of Education. Due to budget cuts in the past, the district has not had a full-time individual responsible for the gifted program in the district since 2011.

Case study #8. Individuals from two of the nine school sites came together to conduct the evaluation of the cluster grouping program jointly. One of the schools is the district's top performing school and the second school is one of the lowest performing. The highest performing school has over 30 students identified as gifted, while the lowest performing school has only five.

Participants. The team consisted of three teachers, one parent, three administrators, and two teacher coaches. Teaching and administrative experience ranged from 12 to 28 years. One teacher taught grade 2, one teacher taught grade 4, and one teacher taught grade 6 during the 2014/2015 school year. The parent had a gifted child enrolled in the district. All evaluation team members were women except for the parent. One teacher was African American and the parent was Hispanic. Five participants reported having no prior experience with program evaluation and the remaining four reported (a) participating in district level improvement plans, (b) evaluating individual

teachers as a part of job responsibilities, (c) being enrolled in a doctoral program of their own, and (d) working in program management in the past.

Pre-study survey results. The evaluation team members from District D agreed that the purpose of evaluation is to identify strengths and weaknesses in order to improve. One participant captured the consensus of the group best when writing the phrase “reinforcement and refinement” (D.H., pre-study survey, December 2014). Only 10 of the 27 adjectives used to describe the gifted program in District D were positive. Example positive adjectives listed included passionate, welcoming, dedicated, and ‘a diamond in the rough’. Seventeen negative adjectives were listed including underutilized, unknown, basic, under-developed, limited, stagnant, and under-funded.

The mean self-reported rating for the amount of program evaluation knowledge possessed by the group on a scale from 1 (Novice) to 4 (Expert) was 1.6 (0.88). This placed the team between the novice and proficient categories. The mean rating of pre-existing knowledge of the Kaplan Depth and Complexity Model on a scale from 1 (I’ve never heard of it) to 4 (I use it all the time) was 2.9 (0.93). This placed the team just below the “I’ve used it a few times in the past” category.

One individual reported having no anxieties and concerns regarding participating in the study. Four team members listed time as the biggest concern, two worried whether or not the district would follow through with supporting the results of the evaluation, and two expressed self-doubt regarding their own ability to follow-through and complete tasks required of study participants.

Experience while using the DC-PET. The evaluation team from District D met a total of 10 times between November and May. I attended four of the meetings in person

at the request of the team leader to support and monitor the process. The team members collaborated using Google Docs and email between meetings. The group reported spending between 17 to 18 hours completing the process and finished 10 ½ of the 11 steps. Recommendations were generated and the results communicated to the board, but the creation of SMART goals was postponed due to the first recommendation (see the *Description of the Evaluation Conducted* section that follows for a more detailed explanation). During this time, the team completed 16 status checks. The content of the questions included requests for additional information about resources available to improve current practices and logistical questions, such as “What parts do I do on my own and what parts do we do as a team?”

I provided feedback and suggestions for modification to the team regarding the wording of the evaluation questions that the team developed. My suggestions were later adopted. An example of one original question was, *What should a stellar gifted program look like and sound like?* I explained that although this was an excellent question, it did not pertain to District D’s gifted program. The final evaluation question that resulted after the team considered my feedback was, *To what degree does District D’s gifted program align with the NAGC standards?* No conflicts were reported between the team members. The team stated that everyone really wanted to be there and was invested in the process. Unlike other teams, every member of the team reported relying more heavily on the online application than the paper workbook.

Description of the evaluation conducted by the team. The evaluation team members decided the purpose of their evaluation would be to measure the effectiveness

of the program and uncover any contradictory practices. The final three evaluation questions developed by the team were:

- To what degree does the District D gifted program align to the NAGC standards?
- What professional development do District D teachers and leaders need in order to align the program to NAGC standards?
- To what extent does the current program ensure that students are being tested and identified properly?

Questions one and two were chosen from the list of unanswered questions generated by the group and question three came from the list of equity and excellence issues.

In order to answer their evaluation questions, the team decided to (a) organize two focus groups, (b) send out a teacher survey, and (c) conduct a document review. The first focus group was conducted with 12 teachers and one administrator in an attempt to answer evaluation question number two. Common themes that arose included (a) significant levels of teacher self-doubt, (b) a perceived lack of training, (c) concern about bored students, and (d) a desire to incorporate more choice into the classroom.

The second focus group was conducted with the counselors in the district in order to help answer evaluation question number three, due to the fact that the counselors are the individuals responsible for identifying gifted students. The big ideas generated from the focus group were: (a) the referral process and deadlines need to be publicized to a greater extent; (b) a uniform process for testing students needs to be developed and followed; (c) communication of the testing results needs to be improved;

and (d) teachers need to receive more training on what to look for when deciding to refer a student for testing.

The survey was distributed to the gifted cluster teachers at the two school sites involved in the evaluation. The team chose to answer evaluation question number one by converting the “Master Checklist of Gifted Program Elements for Self Assessment” document (Speirs Neumeister & Burney, 2012) into an online form. A copy of this checklist has been reprinted with permission from Prufrock Press in Appendix E. Respondents were asked to read 41 statements and decide if there was no evidence of this taking place in District D, some evidence, or if it was firmly in place. Participants were allowed to comment on each statement as well. Seven gifted cluster teachers responded to the survey. The results indicated no evidence found within District D to support the practice for a majority of the statements. The respondents rated the statement *Students are identified in all grade levels for which services are provided* the highest and *Policies are in place to allow early entrance, grade skipping, subject skipping, early credit, and early graduation according to individual student need* the lowest.

A document review of longitudinal testing records and student files was conducted in order to uncover any disparities in the number of students identified in the district by ethnicity and gender. It was revealed that 3% of the district was identified as gifted and the largest disparities existed between (a) the number of boys versus girls, and (b) the number of Whites versus Hispanics. In District D, 57% of students identified as gifted are boys and 43% are girls. In addition, Whites make up 23% of the district, but 43% of the gifted program. Hispanics make up approximately 61% of the district, but only 40% of the gifted program.

The team members generated three recommendations after analyzing the data.

The recommendations were:

- Continue the DC-PET next year, but at all school sites instead of just two.
- Provide training for the counselors on giving the gifted assessment.
- Provide support to teachers describing who to refer for testing (i.e., characteristics of gifted students).

The team, which included District D's Director of Special Education, was uncomfortable creating SMART goals until data collection went district wide. The team voted to continue meeting next year and to recruit at least one individual from each of the nine schools to participate. The results of the evaluation and the plan to continue collecting data next year were presented at a school board meeting.

Evaluating the DC-PET. The response rates for the surveys and focus group can be found in Table 37. The mean rating for evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.6 (0.88) on the pre-survey, but 2.14 (0.69) on the post-survey, indicating a jump from novice to proficient. The mean rating of expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 2.9 (0.93) on the pre-survey, but 3.14 (0.38) on the post-survey. This placed the time in the "I have used it a few times in the past" category.

The mean ratings for the degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at all) to 4 (A Lot) can be found in Table 38. Seven of the 10 principles were given the highest score of 4.0 (0.0) across the board. The lowest rated principle was inclusion, rated slightly lower at 3.71 (0.76).

Table 37

Survey and Focus Group Response Rates for Case Study #8

	<i>N</i>
Pre-Survey Only	2
Post-Survey Only	0
Matched Pre- and Post-Surveys	7
Focus Group Participants	6
Total	9

Note. Out of 9 team members

Table 38

Empowerment Evaluation Principles Ratings for Case Study #8

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	7	4.00	0.00
Inclusion	7	3.71	0.76
Democratic Participation	7	3.86	0.38
Community Knowledge	7	4.00	0.00
Evidence-based Practices	7	4.00	0.00
Accountability	7	3.86	0.38
Improvement	7	4.00	0.00
Organizational Learning	7	4.00	0.00
Social Justice	7	4.00	0.00
Capacity Building	7	4.00	0.00

Note. Response scale ranged from 1 (Not at All) to 4 (A Lot)

The team listed multiple skills learned or refined as a result of using the DC-PET. The first was the importance of listening to others. They also now feel confident conducting a focus group and are open to using other forms of data collection beyond surveys alone. One team member believed her fear of statistics and data analysis decreased as a result of this experience. Y.K. stated:

That was something I really learned from this process. Normally, I get sweaty when I see numbers. I don't like them. But when she started writing those numbers on the board, I actually was like, "Really?" It wasn't just the understanding part, it was more like, "Oh my gosh!" That's really interesting. (Y.K., focus group, May 25, 2015)

A second member enthusiastically stated that she developed a new love for qualitative data analysis.

The group as a whole believed the DC-PET required them to slow down and be more reflective instead of reactive. One team member said:

One of the things that came out from taking the time to do it was that we were able to bring back really solid and rich data. So, it's one thing to say 'I think that this is broken' or 'I think this is not working'. It's very different to be able to say, 'Here's a percentage of kids that we have'. Like this is the real number. Guess what? This is what people believe about how to identify kids. Guess what? This is what people are saying about whatever. (Y.K., focus group, May 25, 2015)

A second person stated, "I think the power behind meeting like this is that I feel there's validity in the process. It's not just what we think or our feelings. There's data, there's reflection time" (W.K., focus group, May 20, 2015).

Team members expressed the belief that when comparing the DC-PET to other forms of evaluation, the DC-PET was (a) self-contained in one nice package, (b) showed the whole picture, (c) used multiple sources, and (d) provided guides when needed. They appreciated the videos, links, and audio recordings embedded in the app and strongly

believed the Kaplan Model icons were beneficial and did not feel like an add-on. For example, D.H. said:

I have a deeper appreciation for the Kaplan Model because of this. How many times have we seen kids using it with English language arts and now, to be like, oh, you can look at it with data. You can look at it with a teacher meeting. You can look at that with, you know, a board meeting. (D.H., focus group, May 25, 2015)

The one suggestion made for improving the DC-PET was finding a way to allow users to record answers to each step electronically online.

Thinking back, the team believed the tool provided just the right amount of structure. One individual felt it was too rigid at first, but changed her views once the evaluation proceeded. She expressed a desire in the beginning to jump right to the recommendations for change section, but now sees the value in data collection and analysis. She commented, “Don’t skip the data collection. It’s very tempting because you feel you know what needs to be done. And maybe you’re right, but it’s just important to hear everything first” (R.M., focus group, May 25, 2015).

The team members also reflected on their own practices and believed next time they would plan a year’s worth of meetings in advance instead of planning week to week. They also believed that the minimum time any one single meeting should last is 90 minutes. The mean rating for how likely they would be to use the DC-PET again in the future on scale from 1 (very unlikely) to 7 (very likely) was 6.57 (0.54). This indicated the team was between “likely” and “very likely” to use the tool again at some point in the future.

District E Case Study

The ninth case study focused on a tribal community school in District E. This rural community is a closed Ojibwee reservation in Minnesota operated by the Bureau of Indian Education. The entire student population is Native American and most have financial need.

Case study #9. The gifted program within the school is limited to a pullout program, along with differentiation in the regular classroom. The teacher of the program started the gifted program when he began teaching in the district several years prior. Previously, no services were provided, and students were not identified.

Participants. The evaluation team consisted of one administrator, two teachers, the school social worker, and an academic coach. One of the individuals was Native American and the remaining four were White. Two participants were men and three were women. Years of teaching and administration experience ranged from 3 to 30 years. One teacher taught grade 4 and the second teacher taught the gifted pull-out classes during the 2014/2015 school year. Four of the team members reported having little or no experience with program evaluation, but the fifth indicated using the Bureau of Indian Education's guidelines to examine the elements of the current gifted program in the past.

Pre-study survey results. The members of the District E evaluation team agreed that the purpose of program evaluation is to determine what is working and what is not working in order to make improvements. Fourteen of the adjectives listed by the group to describe the gifted program in District E were positive. Several examples included

creative, challenging, engaging, socially meaningful, interesting, and growing. The one negative adjective phrase used was *too inclusive*.

The mean rating for the amount of existing program evaluation knowledge possessed by the group on a scale from 1 (Novice) to 4 (Expert) was 1.2 (0.45). This placed the team at the novice level. The mean rating of pre-existing Kaplan Depth and Complexity Model knowledge on a scale from 1 (I've never heard of it) to 4 (I use it all the time) was 1.4 (0.55). This placed the team in the "I've never heard of it" category.

One evaluation team member reported having no anxieties or concerns about participating in the study. Three individuals pointed to the time it would take to complete the process in a meaningful way. The last individual expressed a fear of not understanding the process or the dynamics involved in evaluating a program.

Experience while using the DC-PET. District E's evaluation team met 11 times between the months of November and May. The team spent 10 hours to complete steps one through eight, which involved (a) choosing the evaluation questions, (b) creating or choosing the tools necessary to collect the data, (c) collecting the data, and (d) analyzing the data. The team plans to generate SMART goals and develop a communications plan at the start of the next school year.

Overall, the team completed eight status checks, asking only one question. The team asked how to best collect data in order to answer the question they developed regarding defensible differentiation. No conflicts between the team members were reported. In fact, S.N. stated, "It was a dream" (S.N., focus group, June 1, 2015). All team members reported using the online application more heavily than the workbook,

except when reflecting back on what took place at the previous meeting a day or two before the next meeting.

Description of the evaluation conducted by the team. The District E program evaluation team members decided the purpose of their evaluation would be to measure the effectiveness of the program. The final three evaluation questions developed by the team were:

- What is the awareness level and understanding of the gifted and talented program among teachers and staff, students, student families, and within the community?
- What is the mental health profile of the students enrolled within the gifted and talented program?
- What does defensible differentiation look like within the school as a whole?

Questions one and two were chosen from the list of unanswered questions generated by the group and question three came from the list of equity and excellence issues.

The team created two surveys and conducted one document review in order to answer their evaluation questions. The first survey was sent to all parents within the school, as well as to the entire teaching staff. The questions on the survey focused on awareness of the current gifted program. Thirty-four teachers and 10 parents responded. The second survey focused on differentiation and was sent to staff only. The school counselor conducted the document review in order to determine if the level of mental health needs differed between the identified gifted students and their typical peers.

The first survey revealed parents and teachers were aware that a gifted program existed, but did not understand (a) what the program consisted of, (b) how someone was identified for the program, or (c) what common gifted traits were. The second teacher survey measuring the degree to which differentiation was taking place showed 11 teachers would classify themselves as novices with regard to differentiation, 15 self-identified as proficient, two as advanced, and one as a master. Only seven teachers reported using pre-assessment on a regular basis. In response to this data, a workshop on differentiated instruction was provided to the staff in hopes of raising awareness and sharpening skills. The document review conducted by the school counselor revealed the prevalence of significant mental health needs across the board, regardless of whether the student was identified as gifted or not.

Unfortunately, the team ran out of time to complete all 11 steps before the end of the school year. The team is committed to finishing the process this summer or once school resumes in the fall. SMART goals were not created, but at least two general goals were determined:

- Increase awareness of the gifted program within District E.
- Improve teacher knowledge of differentiated instruction.

A third possible goal discussed by the leader of the team was to pilot the Kaplan Depth and Complexity Model with students the following year. The team explained that many issues and distractions outside of their control limited the ability to finish the process within the time frame of the study. The results will be shared with the school board, the staff, and the Bureau of Indian Education once all 11 steps have been completed.

Evaluating the DC-PET. The response rates for the surveys and focus group can be found in Table 39. The mean rating of the evaluation team’s evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.2 (0.45) on the pre-survey, but 2.33 (0.58) on the post-survey. This indicated a change from novice to proficient. The mean rating of expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I’ve never heard of it) to 4 (I use it all the time) was 1.4 (0.55) on the pre-survey, but 2.67 (0.58) on the post-survey.

Table 39

Survey and Focus Group Response Rates for Case Study #9

	<i>N</i>
Pre-Survey Only	2
Post-Survey Only	0
Matched Pre- and Post-Surveys	3
Focus Group Participants	4
Total	5

Note. Out of 5 team members

The mean ratings for the degree to which the evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at All) to 4 (A Lot) can be found in Table 40. All 10 evaluation principles were rated as being evident within the DC-PET at 3.0 or greater. The lowest rated principles were the community ownership, community knowledge, accountability, and social justice principles, all rated at 3.0 (1.00). The highest rated principles were the democratic, improvement, organizational learning, and capacity building principles, all at 3.67 (0.58).

The team members appreciated (a) the group discussions, (b) the opportunity to read research articles, and (c) the ability to dedicate time to focus on the gifted program. Thinking back, the team believed the tool provided just the right amount of structure. G.R. stated, “It met the Goldilock’s criteria” (G.R., focus group, June 1, 2015) referring to the childhood story *Goldilocks and the Three Bears*.

Table 40

Empowerment Evaluation Principles Ratings for Case Study #9

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	3	3.00	1.00
Inclusion	3	3.33	0.58
Democratic Participation	3	3.67	0.58
Community Knowledge	3	3.00	1.00
Evidence-based Practices	3	3.33	0.58
Accountability	3	3.00	1.00
Improvement	3	3.67	0.58
Organizational Learning	3	3.67	0.58
Social Justice	3	3.00	1.00
Capacity Building	3	3.67	0.58

Note. Response scale ranged from 1 (Not at all) to 4 (A Lot)

They believed the process was valuable, systematic, and important. R.N. stated, “The process was as valuable for us as the product. Regularly meeting to discuss these issues was really worthwhile” (R.N., focus group, June 1, 2015). G.R. stated, “I just thought it was really helpful to do this for our program here. I think it shown [sic] a light on some things that we weren’t paying attention to or that we otherwise wouldn’t have noticed” (G.R., focus group, June 1, 2015).

The participants believed the tool could be used in many settings, not just the gifted program, and reportedly learned new data collection strategies. Individuals also stated that when comparing the DC-PET to other forms of evaluation, the DC-PET was (a) more interactive, (b) systematic, and (c) required a multi-disciplinary team of stakeholders to participate. However, G.R. stated, “It’s not 100% linear” (G.R., focus group, June 1, 2015) as he initially thought and reflected on the iterative nature of the tool.

All team members strongly stated that the Kaplan Depth and Complexity Model served as a guide for the evaluation and simplified the process, as well as made it more efficient. M.W. commented, “I really like the thing with the little icons. It guided us to think about things within certain parameters. It just simplified it” (M.W., focus group, June 1, 2015). In fact, all participants expressed an interest in learning more about the Kaplan model and requested additional resources for using the model with students.

The mean rating for how likely the team would be to use the DC-PET again in the future on a scale from 1 (very unlikely) to 7 (very likely) was 6.67 (0.58). This indicated the team was between “likely” and “very likely” to use the tool again at some point in the future.

Cross-Case Analysis

A cross-case analysis involves combining the results of multiple case studies to identify patterns and themes. I have analyzed the data and described trends in: (a) response rates and team composition; (b) anxieties or concerns; (c) implementation and time investment; (d) comparisons to other methods of evaluation; (e) strengths and weaknesses of the DC-PET; (f) usefulness of the Kaplan Model icons; (g) alignment

with empowerment evaluation principles; (h) effectiveness of the technology; and (i) overall participant ratings.

Response rates and team composition. Table 41 shows the response rates for all nine evaluation teams categorized by stakeholder group. A noticeable decline in participation is evident between the administration of the pre and post surveys. A participant's failure to complete the post-study survey did not necessarily mean they did not contribute during the focus group or during work sessions with other evaluation team members periodically throughout the study. The most common reason given for an absence was a lack of time. For example, J.S. wrote, "Life got in the way" (J.S., focus group, May 19, 2015).

The group experiencing the greatest decline was the parent group. Six out of ten parents did not finish the process. A majority of the parents who left the study were from District C. These parents expressed the belief that school employees were not interested in listening to their opinions. Evaluation teams beginning with more than the required five members (i.e., Districts C and D) failed to retain the large numbers over time. District C experienced a 33% decline in participation and District D experienced a 22% decline. It is also interesting to note that the team with the largest number of participants, District C, did not complete a successful evaluation at all. With the exception of District C, a faithful core of 3 to 6 individuals on each team kept the process on track and reported having deep buy-in.

The participation of students on the evaluation teams must also be examined. First, although given the option, only one team chose to include students as active members. Second, the students' contribution was minimal at best. The students did not

speak out during the meetings, failed to complete the post-study survey, and disengaged from the process.

Table 41

Pre to Post Response Rates by Stakeholder Group for All Case Studies

	Pre Study Only	Post Study Only	Matched Pre/Post
Teachers	4	0	24
Parents	6	1	3
Students	2	0	0
Administrators	3	0	6
Other	2	0	4
Total	17	1	37

Note. The category of ‘other’ refers to school counselors, teacher coaches, and community members.

Anxieties or concerns. Four themes were generated from the pre-study survey after a content analysis using axial coding (see Table 42). The first theme was *concern about time*. This refers to the time intensive nature of the DC-PET. Nineteen out of 55 participants in the treatment group expressed this concern. D.A. wrote, “I am concerned about the time involved as I don’t want to be overwhelmed by work outside teaching” (D.A., pre-survey, November 2014). W.K. wrote, “I think time is always a concern. I worry that participants may not be able to participate the whole time” (W.K., pre-survey, December 2014). K.M. wrote, “Spring is always a very busy time for me” (K.L., pre-survey, January 2015).

The second theme was *nervousness about newness*. Eight out of 55 individuals were anxious about learning a new skill or method of evaluation. H.P. wrote, “I’m not sure what to expect” (H.P., pre-survey, November 2014). K.L. wrote, “No anxieties, just concern over making sure I use the tool correctly so the information received from it is useful” (K.L., pre-survey, December 2014). R.S. wrote, “The only anxiety is one often affiliated with learning something new” (R.S., pre-survey, January 2015).

The third theme was *fear that the results would not be used*. Five participants were afraid that the time and effort spent collecting data would be wasted because the data might not be used. P.H. wrote, “My hope is that if all this time is being invested and recommendations are made, someone please listen and allow changes to be made if need be” (P.H., pre-survey, January 2015). D. H. wrote, “What will happen next in our district once the information is gathered? Will the process to rebuilding move rapidly or slowly?” (D.H., pre-survey, December 2014). M.C. wrote, “I hope that what we learn isn’t purely for academic use, but will also be applied in ways which strengthen our gifted program” (M.C., pre-survey, January 2015).

The fourth theme was *self-doubt of the participants* in their ability to conduct an evaluation. Four participants doubted themselves enough to record this on the survey. M.B. wrote, “It is hard to critically reflect upon myself” (M.B., pre-survey, November 2014). Y.K. wrote, “My inexperience is a concern” (Y.K., pre-survey, December 2014). Similarly, M.W. wrote, “not understanding all of the dynamics that are needed” (M.W., pre-survey, November, 2014). Lastly, L.C. wrote, “Lack of experience in the process and being able to be an effective, contributing member of the team” (L.C., pre-survey, January 2015).

Table 42

Frequency of Anxieties or Concerns About Participating Themes

Themes	Number of Participants	Frequency of Occurrence
Concern about time	55	19
Nervousness about learning something new	55	8
Fear that the results would not be used	55	5
Self-doubt of the participants	55	4

Implementation and time investment. Each group differed in how they chose to implement the DC-PET process. The districts with more than one evaluation team involved in the study took two different approaches. District B broke up into three smaller teams, but chose to meet together at the same time. District A, however, broke up into three smaller teams, but worked completely independently of one another with no communication across teams. Also, the gifted coordinator for District B was actively involved in the evaluation process from start to finish, but the gifted coordinator from District A did not participate until the teams finished at least step nine.

Districts A and B chose to look at one particular gifted program service, such as cluster grouping or self-contained gifted across many schools, while Districts C and E examined how the program was being implemented only at one particular school. The remaining team, District D, was composed of team members from two schools that both offered the same service model. The team members from both schools worked together as one, helping each other along the way to complete one joint evaluation.

The amount of time evaluation teams reported spending on the process ranged from 10 - 18 hours per team member (see Table 43). It is important to note that the total time invested in the process for case studies 7 – 9 would have been greater if they had completed all 11 steps within the six months time frame. Eight of the nine teams reported using some form of electronic communication and collaboration between meetings. Steps five and seven on average took the most time. Step five involved choosing the final evaluation questions and step seven involved deciding who was responsible for each step in the evaluation, as well as creating all necessary instruments such as surveys or focus group protocols. Seven out of nine teams agreed that the process of writing evaluation questions was the most purposeful, but challenging part of the entire process. P.H. wrote, “It was hard to come up with questions that would lead to quantifiable data” (T.H., post-survey, April 2015). J.M. wrote, “I thought it was interesting trying to get to the very best question we could to collect the information that we wanted. I really learned a lot from that” (J.M., focus group, April 22, 2015). M.C. wrote, “I like that as we tried to develop questions, issues of concern regarding the program were raised” (M.C., post-survey, April 2015).

Eight of the nine evaluation teams conducted successful evaluations using the DC-PET. For the purpose of this study, a successful evaluation using the DC-PET was defined as one in which the participants completed step eight or greater (i.e. created evaluation plan, collected data, identified patterns and trends) and expressed an intention to finish the remaining steps the following semester. Evaluation teams one through six completed all 11 steps. Evaluation team seven stopped at step five due to a split that arose between the parents and school staff. Evaluation team eight completed steps one

through nine and evaluation team nine stopped at stop eight. Both teams plan to finish the process during the next semester. Four teams relied solely on surveys to collect data and four used surveys combined with focus groups, document review, needs assessments, or interviews.

Table 43

Average Per Person Self-Reported Number of Minutes Engaged

	Team #1 (n=4)	Team #2 (n=5)	Team #3 (n=4)	Teams #4 - #6 (n=8)	Team #7 (n=7)	Team #8 (n=6)	Team #9 (n=4)	M (SD)
Step 1	90	60	30	90	120	60	30	68.6 (33.4)
Step 2	180	60	15	90	30	30	30	62.1 (57.7)
Step 3	30	30	60	60	180	90	30	68.6 (53.9)
Step 4	90	60	60	120	120	150	90	98.6 (33.4)
Step 5	60	90	60	60	360	90	30	107.1 (113.4)
Step 6	60	90	15	60	N/A	90	30	57.5 (30.6)
Step 7	120	120	240	210	N/A	240	210	190.0 (55.9)
Step 8	90	90	60	60	N/A	120	120	90.0 (26.8)
Step 9	60	60	60	60	N/A	60	N/A	60.0 (0.0)
Step 10	N/A	60	60	60	N/A	60	N/A	60.0 (0.0)
Step 11	N/A	30	60	60	N/A	120	N/A	67.5 (37.7)
Per Person	780	750	720	930	810	1,110	570	810 (170.6)
Per Team	3,120	3,750	2,880	2,480	5,670	6,660	2,280	3,834 (1,684.9)

Comparisons to other methods. The first qualitative question was: *How did the DC-PET compare to previous methods of program evaluation?* The feedback from the post-surveys was overwhelmingly positive. Only the responses of those who indicated previous experience with evaluation were counted ($n=22$). In all, 19 out of 22 individuals listed one or more ways in which the DC-PET was better than other methods of evaluation of which they were aware. One individual reported she preferred other methods.

Three major features were identified regarding how the DC-PET compared with other evaluation models using axial coding (see Table 44). First, the DC-PET was *more comprehensive than* other methods of evaluation. Eight individuals pointed to this as a major difference. D.H. wrote that she wished other evaluations showed the “whole piece like this one did” (D.H., post-survey, May, 2015). S.B. wrote, “The DC-PET process had more of a whole or big picture theory” (S.B., post-survey, May 2015). M.C. wrote that the DC-PET “used some similar processes, but broadened our perspective” (M.C., post-survey, May 2015). A.L. noted that it was “very in-depth” (A.L., post-survey, May 2015). J.S. echoed that statement by saying, “It was more thorough and also a more lengthy, involved process” (J.S., post-survey, April 2015).

The second feature, identified by seven individuals, was that the DC-PET was *more organized and structured*. H. P. wrote, “It seemed more structured and took you through the process smoothly” (H.P., post-survey, April 2015). G.R. wrote, “This was more systematic, thoughtful, and –importantly- involved a multi-disciplinary team; because of that, I feel that the evaluation is more meaningful” (G.R., post-survey, May, 2015). G.B. wrote, it was similar to other approaches, but she “enjoyed the interactive

and systematic approach” (G.B., post-survey, May 2015). K.L. wrote, “This process was much more intentional and structured specifically for gifted programming” (K.L., post-survey, May, 2015).

The third feature, identified by four individuals was that the DC-PET was *clear and easy to use*. A.W. wrote, “I am a novice, but the tool was very easy to use and logical to follow” (A.W., post-survey, April 2015). K.S. wrote that it was “all easily contained in one location” (K.S., post-survey, May 2015). C.G. wrote that the DC-PET was “very well organized, detailed, step-by-step, easy to follow and understand” (C.G., post-survey, May, 2015). W.K. wrote, “It was helpful to have a clear process to follow with guides when needed” (W.K., post-survey, May, 2015).

Table 44

Frequency of Themes Related to How the DC-PET Compares to Other Methods

Themes	Number of Participants	Frequency of Occurrence
More comprehensive	22	8
More organized and structured	22	7
More clear and easy to use	22	4

Strengths and weaknesses. When participants were asked to describe what they liked best about the DC-PET, five categories emerged through axial coding (see Table 45). The most frequently mentioned category was the ability to *collaborate and discuss with peers*. Seventeen of the 38 participants who completed the post-survey mentioned

this aspect of the process. T.H. wrote, “It brought up good discussion” (T.H., post-survey, April 2015). J.S. wrote, “The tool got our conversations focused and going in the right directions” (J.S., post-survey, April 2015). G.R. wrote, “What I liked best, and also felt was the most valuable, was the time spent together with the other team members discussing our program” (G.R., post-survey, May, 2015).

Table 45

Frequency of Themes Related to What Participants Liked Best About the DC-PET

Themes	Number of Participants	Frequency of Occurrence
Collaborate and discuss with peers	38	17
See other’s perspectives	38	13
Feedback based on data	38	12
Many useful resources	38	8
Learning of new skills	38	8

The second most frequently mentioned category was that the DC-PET allowed the evaluation teams to *see other’s perspectives*. Thirteen individuals appreciated the ability to capture other stakeholders’ opinions. Y.K. wrote, “I loved the discussion it inspired. It helped to put the good, the bad, and the ugly right out on the table for us all to see and face head on” (Y.K., post-survey, May, 2015). P.H. wrote, “The instrument forced me to be open minded” (P.H., post-survey, April 2015). C.Y. wrote that she

“liked breaking down the program and seeing the parents’ point of view along with the teachers” (C.Y., post-survey, April 2015). D.H. wrote:

The greatest part I would have to say is working side-by-side with district officials who have a different point of view than I in the frontline with students, in order to understand the larger context in which the conversations about gifted happen. (D.H., post-survey, May 2015)

K.L. wrote, “The best part for me, being the administrator, was the intensive discussions with stakeholders in which they shared their multiple perspectives” (K.L., post-survey, May 2015).

The third category, mentioned by 12 of the 38 individuals, was that the DC-PET empowered the evaluation teams to generate *clear feedback based on data*. C.G. noted that she liked “getting results for our program and information that will improve practices” (C.G., post-survey, May, 2015). K.L. wrote:

Working through the steps in the study was a valuable experience. The steps forced a systematic approach to reflecting, questioning, collecting and analyzing data, then setting goals and action steps based on the data. This is not a process I would have been able to do before the study. (K.L., post-survey, May 2015)

W.K. wrote, “It was an open process that forced us not to jump right to solutions, but to really look at the problem, gathering data to make decisions” (W.K., post-survey, May 2015). D.A. wrote, “The feedback was clear and lead to SMART goals that were data driven and focused” (D.A., post-survey, April 2015).

The fourth and fifth most often mentioned favorite aspects of the DC-PET were the fact that it contained *many useful resources* and facilitated the *learning of new skills*.

Eight individuals mentioned the fourth and fifth categories. C.G. wrote that she “Loved learning about the Kaplan Model as well as becoming more knowledgeable on how to evaluate a program” (C.G., post-survey, May 2015). J.S. attributed “participating in survey creation, distribution and evaluation of results” while using the DC-PET to the gains she made in evaluation knowledge pre to post (J.S., post-survey, May 2015). G.B. enjoyed “reading the research papers and considering the implications in the research as it applies to our school” (G.B., post-survey, May 2015). C.C. wrote, “Every step was clearly outlined and included all of the resources needed to effectively evaluate the program” (C.C., post-survey, May 2015).

In response to the question *What recommendations for change would you recommend for the DC-PET*, six respondents left the question blank. Twelve of the 38 participants wrote there is nothing they would change. For example, R.C. wrote, “None- the DC-PET worked really well for our group” (R.C., post-survey, April 2015). D.H. wrote, “Get it in the hands of all we know” (D.H., post-survey, May, 2015). Y.K. wrote, “None. I thought it was a fantastic process. I believe every step was beneficial” (Y.K., post-survey, May 2015).

Three individuals wanted to see a new introduction created that was more detailed and informative. D.A. wrote, “I found the initial introduction to what we were supposed to be doing somewhat confusing” (D.A., post-survey, April 2015). K.P. wrote, “Give the big picture at the start. It helps me if I know where I am going before I start” (K.P., post-survey, May 2015). M.C. wrote, “Outline the goals and objectives more clearly to ensure that all have similar expectations” (M.C., post-survey, April 2015).

Three other individuals suggested creating additional management tools teams could use during the evaluation such as premade, but customizable, meeting agendas or a template for the final report. K.L. wrote:

The content and process of the evaluation tool itself are wonderful and gave us the results that I need. I simply needed some management tools since I am an administrator for 29 schools with no assistant and no budget. (K.L., post-survey, May 2015)

L.B. wrote, “I think there needs to be more support structures built into the program” (L.B., post-survey, April 2015).

Other suggestions mentioned by single individuals included: (a) add more space to write in the workbook; (b) provide a more defined timeline; and (c) provide more instruction in the data collection and analysis sections of the tool. Regarding more space in the workbook, C.J. wrote, “Add more writing and recording areas for notes and thoughts. Possibly make the booklet online and interactive” (C.J., post-survey, April, 2015). G.R. spoke about the timeline when he suggested, “a more defined time-line of when each step should be completed” needs to be provided (G.R., post-survey, May, 2015). J.S. wrote, “Teams need more support up front in knowing about types of data that can be gathered, as well as how to gather the data” (J.S., post-survey, April 2015).

Usefulness of the Kaplan model icons. The second qualitative question was: *To what extent does the Kaplan Depth and Complexity Model provide a useful framework for conducting program Evaluations?* The answer to this question seemed to vary by team. Four teams were enthusiastically supportive of the icons being used in this way and felt they (a) served as road signs for the process, (b) helped make the evaluation

more efficient, and (c) served as a solid framework. No real patterns could be identified among the supporters of the icons due to the fact that two were from high performing District A and two were from lower performing Districts D and E. Prior experience with the Kaplan Model was not consistent among those who saw the benefit in the icons either. Ratings for prior experience among supporters of the Kaplan Model ranged from 1.4 (0.55) to 2.9 (0.93) on a scale from 1 (I've never heard of it) to 4 (I use it all the time).

The teacher participants from District C saw the benefit of using the icons, but the parents on the team did not. Only three out of the nine participants in District B's focus group reportedly used the icons during the evaluation. One individual stated the pictures designed to represent each icon did not match the meaning and were therefore confusing. Only one person on the third team from District A reportedly found the icons helpful. The one person who did use them, stated that she uses the icons in her classroom, and the application of the model in this new way worked very well.

On the positive side, the evaluation team from District E became interested in the model and plans to possibly teach the icons to students next year. At least two participants from District A also plan to learn more about the Kaplan Model. Although many individuals did not use the icons, there is no evidence to suggest that including them in the DC-PET created any barriers or did any harm. The individuals not using the icons simply ignored them, and proceeded to the details of the task that icon introduced. It is also important to note that all teams grew in their self-reported understanding of the Kaplan Model despite their opinions as to its value. Surprisingly, one individual from District A reported never hearing of the Kaplan Depth and Complexity Model at the

conclusion of the study. This could possibly be explained by confusion in reference to the words “Kaplan model” versus the term “ depth and complexity icons” used in the app.

Alignment with empowerment evaluation principles. The third qualitative question was: *To what degree did the DC-PET align with the 10 principles associated with empowerment evaluation?* The participants in the study believed that all 10 empowerment principles are present in the DC-PET frequently or very frequently as assessed by the post-study survey. The mean ratings for the degree to which all evaluation team members believed the DC-PET aligned with the 10 empowerment evaluation principles on a scale from 1 (Not at All) to 4 (A Lot) can be found in Table 46. All 10 principles were rated 3.11 or greater. The average of all ratings was 3.34 (0.13).

Table 46

Empowerment Evaluation Principles Ratings for all Case Studies #1 - #9

Principle	<i>N</i>	<i>M</i>	<i>SD</i>
Community Ownership	38	3.37	0.68
Inclusion	38	3.37	0.79
Democratic Participation	38	3.47	0.69
Community Knowledge	38	3.29	0.73
Evidence-based Practices	38	3.50	0.65
Accountability	38	3.21	0.88
Improvement	38	3.50	0.76
Organizational Learning	38	3.26	0.89
Social Justice	38	3.11	0.95
Capacity Building	38	3.29	0.84

Note. Response scale ranged from 1 (Not at All) to 4 (A Lot)

The lowest rated principle was the social-justice principle at 3.11 (0.95). When analyzing the equity and excellence issues chosen by each evaluation team, it is not surprising that this principle was rated the lowest, although still rated at the ‘frequently’ mark. Four teams chose defensible differentiation and four teams chose affective needs of students. Only one team decided to examine patterns in identification in order to look for underserved populations. The topic of underserved populations is the most obviously social justice oriented of the four choices from which the teams could choose. Also, the four teams that chose to examine defensible differentiation did not truly examine whether or not *defensible* differentiation was taking place, but simply focused on whether differentiation in general was taking place. The central argument of Borland’s (2009) defensible differentiation concept is that schools should not keep non-gifted students from being exposed to skills often reserved for identified gifted students (e.g., critical thinking skills training, creative problem solving, Socratic seminars). This aspect of his theory, if applied in an evaluation, would more concretely and obviously demonstrate social justice in action.

The highest rated principles were evidence-based practices and improvement at 3.50. Several teams commented at the conclusion of the evaluations that it was very nice to have actual data to justify making changes to the program instead of simply relying on a gut feeling or accepted belief. Although four teams chose to rely completely on survey data that focused on perceptions of stakeholder groups, four attempted to triangulate their data by combining survey data with a focus group, document review, or interview. It is important to “meet programs where they are at” (Fetterman & Wandersman, 2007, p. 37) and celebrate the fact data collection was undertaken at all.

Perhaps during a future evaluation these teams may venture into examining student outcomes or other more complicated procedures.

Regarding the improvement principle, every member of every team initially expressed an interest in engaging in a process that would help to improve their gifted program. Although a percentage of participants chose to disengage from the process at some point as the evaluation proceeded, their initial motivation was to see the program grow and change. Financial gain did not seem to make a difference in the quality of the evaluation that resulted from each teams' evaluation either. Only three of the eight teams that completed a successful evaluation received money for their time. Each team, regardless of the size they began with, ended up having a core group of 3 to 5 individuals who were truly invested in the evaluation. Due to this fact, as a general rule, I would advise all new evaluation teams to stay within the five-person limit.

Effectiveness of technology. The fourth qualitative question was: *How did the technology component of the DC-PET affect the experience of using the tool?* There was unanimous agreement that the online application and the paper workbook were useful. The majority of individuals believed the online application was most useful during (a) the beginning of the evaluation and (b) the data collection phase. The participants watched the videos and listened to the audio recordings in order to learn about the Kaplan Model and the DC-PET early on and then relied heavily on the resources linked in the app during data collection. The links and attachments in this section gave step-by-step instructions for conducting various data collection techniques and provided sample, validated instruments, such as Gentry and Gable's (2000) *My Class Activities*. Most used

the paper workbook as a refresher before each meeting and to remind themselves where they were in the process.

An analysis of the number of sessions the electronic version of the DC-PET was used, as well as the average length of each session, can be found in Table 47. These figures were generated by Google Analytics, a free online tool offered by Google. Standard deviations are not included due to the fact this information is not available on Google Analytics.

Table 47

Relevant Google Analytics Results for the DC-PET

State	Sessions	Users	Mean Length of Session
Arizona (Districts A-D)	414	139	9.34 min.
Minnesota (District E)	39	11	13.38 min.
Total	453	150	11.36 min.

Overall participant ratings. After the data from all focus groups and post-surveys were combined, the resulting data showed that 51% of the participants on the evaluation teams were very likely or likely to use the DC-PET again in the future. A breakdown of all possible responses can be found in Table 48. The mean rating on a scale from 1 (very unlikely) to 7 (very likely) was 5.45 (1.31).

The mean rating of team members' completing both the pre and post-surveys for evaluation knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.46 (0.61) on the pre-survey, but 2.19 (0.57) on the post-survey. This indicated a jump from novice to

proficient. The mean self-rating of all participants' completing both the pre and post-surveys for expertise in using the Kaplan Depth and Complexity Model on a scale from 1 (I have never heard of it) to 4 (I use it all the time) was 2.24 (1.03) on the pre-survey, but 2.97 (0.74) on the post-survey. Tables 49 and 50 provide a breakdown for the percentage of respondents choosing each possible answer

Table 48

Likelihood of Using the DC-PET Again in the Future / Case Studies #1 - #9 (n=38)

	<i>N</i>	Percentage
Very Unlikely	0	0%
Unlikely	1	3%
Somewhat Unlikely	2	5%
Undecided	5	13%
Somewhat Likely	11	29%
Likely	9	24%
Very Likely	10	27%

Table 49

Self-Rating of Program Evaluation Knowledge for Matched Pre and Post

	Pre-Study (<i>n</i> =37)	Post-Study (<i>n</i> =37)
Novice	60%	8%
Proficient	35%	65%
Advanced	5%	27%
Expert	0%	0%

Table 50

Self-Rating of Kaplan Model Knowledge for Matched Pre and Post

	Pre-Study (n=37)	Post-Study (n=37)
Never heard of it.	30%	3%
Know what it, but have never used it.	30%	19%
I have used it a few times in the past	27%	57%
I use it all the time.	13%	21%

Mixed-Methods Results

All quantitative and qualitative data were analyzed separately and then combined. Pattern matching, member checks, and the cross-case analysis were used to answer the mixed methods questions.

Mixed-Methods Question #1

The first mixed-methods question was: *How do the qualitative results explain or expand on the quasi-experimental outcomes?* Qualitative data collected during the focus groups and post-survey at the conclusion of the study supported the statistical findings already discussed. Discovering that the tool did indeed lead to a significant difference between the treatment and comparison group represented the “what”, but the qualitative data collected during the focus groups, status checks, and surveys represent the “how”. Most notable is the fact that the mean rating for all evaluation teams’ knowledge of program evaluation increased from the beginning of the study to the end. The mean rating of team members’ completing both the pre and post-surveys for evaluation

knowledge on a scale from 1 (Novice) to 4 (Expert) was 1.46 (0.61) on the pre-survey, but 2.19 (0.57) on the post-survey. This indicated a change from novice to proficient.

Each team described and outlined the reasons for the growth and listed new skills they acquired as they engaged with the DC-PET. Despite early concerns regarding the time it would take and self-doubt about their ability to meaningfully participate, most individuals finished the process and discovered it was not as hard as it first seemed. A question on the post-study survey stated, *What has lead to any changes, if any, in the level of expertise you indicated above as compared to the pre-study rating?* Typical responses attributed growth to using the steps within the DC-PET, engaging with knowledgeable co-workers while using the DC-PET, and having the opportunity to participate on an evaluation team for the first time. Although several individuals left this question blank, no one attributed gains in evaluation knowledge to anything other than being a part of the evaluation team while using the DC-PET.

Many individuals remarked that they developed a newfound respect for evaluation and learned numerous skills from participating in the DC-PET process. In addition, the mean rating for the likelihood of participants to use the DC-PET again in the future was 5.45 (1.31) on a scale ranging from 1 (very unlikely) to 7 (very likely). This indicated a strong likelihood that the participants will engage in evaluation again using the tool sometime in the future. This is important to note considering at least a third of the individuals reported never evaluating a program before.

Mixed-Methods Question #2

The second mixed-methods question was: *What modifications should be made to the DC-PET in the future based on both the quantitative results and the findings*

generated from the qualitative data? Based on participants' feedback, three key changes will be made within the DC-PET in the future, as well as several minor modifications. The first key change will be to create a more detailed introduction to the DC-PET. This introduction will include a: (a) description of the empowerment evaluation principles and how they form the bones of the tool; (b) description of the components of evaluative thinking; (c) readiness checklist for allowing users to self-assess whether or not the tool will meet their needs; and (d) list of frequently asked questions.

A second key change will be to investigate and incorporate a way to enable the online application to support electronic data entry. There are many free stand-alone resources already available, but it would ultimately feel more streamlined to have the capability to collaborate online built into the application. I will create additional writing space within the paper workbook as well.

The third key change will be to create and include additional management tools for the leaders of evaluation teams using the DC-PET. Possible tools include (a) pre-made customizable meeting agendas, (b) sample evaluation reports, and (c) recruitment tools for soliciting volunteers to participate on the evaluation team. I will also add additional support and resources to the section of the online application that deals with creating and selecting the evaluation questions. A two part litmus test will be included that instructs the participants to (a) ask themselves if the questions they developed pertain to their gifted program, and (b) ask themselves if the wording of the question is such that the need for change is not assumed. For example, the question "What does a quality gifted program look like?" or "How can we improve the gifted program?" would

be modified to read, “How does our program compare to the national gifted standards?” Numerous examples of quality evaluation questions will be included as well.

Possible minor modifications to the DC-PET include (a) switching the order of the two parts contained within step one, (b) renaming the *equity and excellence issues* as *social justice issues*, and (c) refining the links included. Making these few modifications will hopefully result in a more user-friendly tool.

CHAPTER 5. DISCUSSION

Introduction

The importance of program evaluation to gifted education cannot be overstated. Despite this, many barriers have kept school districts from investing time and money into this practice. The purpose of this mixed-methods study was to pilot and examine the effectiveness of a new program evaluation tool designed to help alleviate some of those barriers called The *Depth and Complexity Program Evaluation Tool* (DC-PET). The DC-PET is different from all other methods previously used to evaluate gifted programs in that it incorporates: (a) the Kaplan Depth and Complexity Model; (b) the 10 empowerment evaluation principles; and (c) appreciative inquiry.

Over the course of six months, nine evaluation teams representing five different districts in two states implemented the DC-PET for the first time. Eight of the nine teams successfully completed an evaluation of a gifted program. Quantitative data were gathered from 55 treatment group participants and 40 comparison group participants in pre/post fashion using the Evaluative Thinking Inventory (Buckley & Archibald, 2011). Exploratory factor analysis, confirmatory factor analysis, and repeated measures ANOVA were used to analyze the data. Qualitative data were collected from the 55 treatment group participants before, during, and after the study. Nine case studies and a

cross-case analysis yielded results. What follows is a discussion and analysis of the results, limitations of the study, and suggestions for future research.

Review and Discussion of Results

Quantitative Results

The quantitative research question explored in this quasi-experimental study was: *To what extent will using the DC-PET result in an increase in the participants' evaluative thinking as measured by the Evaluative Thinking Inventory* (Buckley & Archibald, 2011)? The treatment group and comparison group participants were asked to complete the Evaluative Thinking Inventory created by Buckley and Archibald (2011) from Cornell University's Office for Research on Evaluation in pre/post fashion. Due to the lack of validity and reliability information on the Evaluative Thinking Inventory, exploratory factor analysis was used with the pre-study data and confirmatory factor analysis was used with the post-study data. Descriptive and inferential results were generated by comparing the treatment and comparison group's pre-study Evaluative Thinking Inventory data with the post-study data of both groups using repeated-measures ANOVA.

The analysis of the Evaluative Thinking Inventory provided evidence to support the validity and reliability of the tool to measure the degree to which respondents *believe in and practice evaluation* (factor one) and *pose thoughtful questions and seek alternatives* (factor two). Evidence for the third factor, *describing and illustrating thinking*, was not sufficient to draw conclusions, leaving room for additional work in the future. The next logical step is to collect a larger number of completed inventories and pilot new questions for factor three.

The results of the repeated measures ANOVA revealed no difference between the treatment and comparison groups on the pre-study administration of the Evaluative Thinking Inventory, but a statistical difference on the post-study administration of the instrument. This suggests that using the DC-PET can increase the frequency with which participants think like evaluators. Analysis of within group differences revealed a significant difference for both the treatment and comparison groups.

Finding a significant difference between the treatment and comparison group on the post-study Evaluative Thinking Inventory was not a surprise. I expected to see an increase in participants' knowledge and skills after using the DC-PET. This finding was similar to those found by Robinson, Cotabish, Wood, and O'Tuel (2014) in the Arkansas Evaluation Initiative and Moon (1996) after using the Purdue Three-Stage Model to provide professional development on evaluation techniques.

Finding a significant within group difference between the pre and post administration of the Evaluative Thinking Inventory for the comparison group, however, was a surprise. It is important to note that the difference experienced by the treatment group was much stronger than that of the comparison group, but change occurred nonetheless. This leads me to suspect that repeated exposure to the Evaluative Thinking Inventory itself, even if the respondent has not actively participated in evaluation or sought out information on program evaluation, can lead to a stronger awareness of evaluation and, therefore, a change in the frequency with which one thinks about evaluation.

The eight completed evaluations conducted by the evaluation teams in this study have similarities and differences to those reported in the literature. First, Callahan and

Hunsaker (1993) found a majority of program evaluations in the field of gifted education focused on two concerns: (a) curriculum and instruction; and (b) identification practices and trends. A close look at the evaluation questions developed by the teams in this study mirror those findings.

Second, the fears and anxieties expressed by the participants in this study echoed those of the participants in the Arkansas Evaluation Initiative. Robinson, Cotabish, Wood, and O'Tuel (2014) found that individuals participating in their study were, "moderately concerned about time, logistics, resources, and their skill in facilitating changes needed to improve formative program evaluations in their school" (p.349).

Third, Callahan and Hunsaker (1993) found most evaluations conducted in gifted education relied heavily on questionnaires. In this study, four of the teams used surveys alone and four used surveys combined with focus groups, document review, or interviews. Professional evaluators agree that multiple sources of data provide a more credible picture of the program being evaluated than relying on one alone (Yarbrough, Shulha, Hopson, & Caruthers, 2011).

It is also noteworthy that Moon (1996) reported lower completion rates for the evaluations in her study. Three out of 17 evaluation teams involved in the Purdue Three-Stage Model evaluation institutes succeeded in creating an evaluation plan and implementing it. Eight teams partially completed the process, and four did nothing. In this study, six of nine evaluation teams created an evaluation plan and completed an evaluation along with SMART goals and a communications plan. Two additional teams completed step eight or greater, but only one decided not to continue. This might have been the case because the DC-PET provided the information participants needed to

complete the evaluation prior to their engagement with the task. Moon (1996) mentioned that the evaluations in her study were not assigned as a culminating activity at the conclusion of the training, but prior to its completion.

Qualitative Results

The first qualitative question stated, *How did the DC-PET compare to previous methods of program evaluation?* The participants were asked to write a response to this question on the post-survey. Nineteen out of 22 individuals with prior evaluation experience listed one or more ways in which the DC-PET was better than other methods of evaluation and one individual reported they preferred a different method. The most commonly reported benefits to using the DC-PET over other methods were that it is more: (a) comprehensive, (b) organized, and (c) easy to use. Comments from the focus group also supported the findings from the post-study survey. Strengths of the DC-PET listed by the participants included the facilitation of: (a) collaboration and discussion; (b) seeing other's perspectives; (c) feedback based on data; (d) access to many useful resources; and (e) learning new skills.

The strengths of the DC-PET mentioned by the stakeholders align with attributes that define high-quality program evaluations listed and defined by Callahan (2006). The first characteristic listed by Callahan is *responsiveness*, which she defines as taking into consideration the concerns of the stakeholders. The evaluation team members in this study mentioned *seeing other's perspectives* as a benefit to using the tool. A second characteristic listed by Callahan is *fairness and impartiality*, which she defines as giving equal voice to all. The evaluation team members in this study listed *collaboration and discussion* as the number one strength of the DC-PET.

The one negative comment written by a member of District C's evaluation team stated that the DC-PET provided "less leadership" than other methods. Although a program evaluation model or tool cannot in itself provide leadership, that takes human initiative, the point the respondent is making is logical considering the dynamics that took place within District C's evaluation team. As stated in case study #7, several major issues hampered the team's ability to complete an evaluation. First, no clear leader took charge of facilitating the group. Second, a power-struggle developed between the teachers and the parents. The conflict centered on whose perception of the program and future vision was correct. As discussed by Lord, Ross, and Lepper (1979) humans are naturally inclined to belief preservation. However, practicing critical thinking can change this. Instead of attempting to remedy the situation by providing additional program evaluation knowledge, a better approach would have been for me to discuss evaluative thinking and engage the participants in a simulation or directed practice followed by self-reflection.

It could be argued that District C would be better served by a different evaluation approach that does not require multiple stakeholder groups. This would take care of the teacher versus parent problem, but would severely hurt the evaluation by not taking advantage of the community knowledge principle and the democratic participation principle. The vast majority of professional evaluators surveyed by Fleischer and Christie (2009) also agreed that stakeholder involvement is essential.

Fetterman and Wandersman (2005) stated that prior to beginning an empowerment evaluation, interviews or meetings should take place between the evaluator and the participants in order to explain empowerment evaluation and correctly

set expectations between the evaluator and the team. Although I held meetings with the gatekeepers at each site and leaders of each team prior to launching the DC-PET to explain empowerment evaluation and obtain buy-in, it is clear that all evaluation team members should have been included in the decision to use empowerment evaluation. Fetterman and Wandersman (2007) stated that, “No one empowers anyone... Empowerment evaluators help create an environment conducive to the development of empowerment” (p.182). The decision to engage in the process and become empowered is a choice. All stakeholders need to be aware of this fact.

The second qualitative research question was, *To what extent did the Kaplan Depth and Complexity Model provide a useful framework for conducting a program evaluation?* The DC-PET represents the first time the Kaplan Model has been used to conduct an evaluation. This is similar to how Moon (1996) was the first to use the Purdue Three-Stage Model for conducting evaluation institutes designed to help educators create evaluation plans. Data for this question were generated during the focus group. The answer to this question varied by group. Four teams and the teachers on a fifth team felt strongly that the icons served an important purpose. The remaining three teams had both supporters and detractors. No clear patterns emerged. Those not using the icons simply ignored them and completed the requested task. The benefits of using the Kaplan Model icons when conducting an evaluation described by more than half of participants far outweighed the minor inconvenience of overlooking them for those who did not see their value.

As mentioned before, only two research studies prior to this dissertation have been completed exploring the Kaplan Depth and Complexity Model. My results support

those found by Lauer (2010) that the skills the icons represent are needed and practiced in many disciplines. The fact that a majority of the individuals in my study appreciated their value and spoke of their usefulness in an evaluation context provides support for adding program evaluation to the list of disciplines already examined.

Qualitative question three stated, *To what degree did the DC-PET align with the 10 empowerment evaluation principles?* All treatment group participants were asked to rate the degree to which they felt the DC-PET aligned with each empowerment principle on a scale from 1 (Not at All) to 4 (A Lot). The mean rating for all 10 empowerment evaluation principles across teams was 3.34 (0.13), meaning all 10 principles were exhibited frequently. The lowest mean rating was for the social justice principle. The highest rated principles were the evidence-based practices principle and the improvement principle.

All of the participants in this study were new to empowerment evaluation and underwent a short learning curve to understand that the person filling the role of the professional evaluator would be just one voice at the table (Fetterman & Wandersman, 2005). It was necessary for me to remove the facilitator hat and put on the expert hat to provide targeted advice at times, but the team members made all the decisions. The evidence in this study shows that the DC-PET did indeed “foster self-determination and responsibility instead of dependency” (Fetterman & Wandersman, 2005, p. 32), which is a major goal of empowerment evaluation.

The program evaluation standards created by the Joint Committee on Standards for Educational Evaluation and the evaluation standards included in the National Association for Gifted Children standards call for an expert in evaluation to guide the

process. This is still important when empowerment evaluation is used. While I was not physically present during a majority of the evaluation team meetings conducted by the nine teams in this study, I still provided support and guidance through email, status checks, and in-person feedback when requested. For that reason, it would be important for an individual knowledgeable in program evaluation to facilitate the DC-PET process or be available for support if the team members need guidance.

One of the most identifiable features of empowerment evaluation is capacity building (Fetterman & Wandersman, 2005). Evidence of the capacity building principle collected through this study provides qualitative and quantitative evidence to show that participants did indeed learn new skills and think more often like an evaluator. Twenty-one out of 33 focus group participants were able to list a new skill they learned from participating in the DC-PET process. Furthermore, at least eight of the nine teams now know how to complete a program evaluation despite the fact that one third of the participants had never done so before.

A second identifiable feature of empowerment evaluation is the attention to social justice issues. As a result of this study, only District D answered Yoon and Gentry's (2009) call to closely examine the representation of underserved populations in gifted programs. Four teams focused on the social and emotional needs of gifted students and succeeded in collecting data regarding the issue. The progress made by the remaining teams at addressing social justice was minimal. More work needs to be done in this area. Robinson, Cotabish, Wood, and O'Tuel (2014) also had limited success in affecting the underrepresentation of minority students in gifted programs. The

nomination of culturally diverse students increased, but the identification rate did not ultimately change.

The fourth qualitative question stated, *How did the technology component of the DC-PET affect the experience of using the tool?* Participants were asked if they preferred the paper workbook or the online application, and Google Analytics were obtained showing the degree to which the online application was used. There was unanimous agreement that both the online application and the paper workbook were useful. The majority of individuals believed the online application was most useful during (a) the beginning of the evaluation, and (b) the data collection phase. Google Analytics calculated that the DC-PET was used 453 times during the study and the mean session length was 11.36 minutes.

The DC-PET app is the only technology based evaluation tool currently available in gifted education. The results of this study show that professional development on program evaluation can be effectively provided to stakeholders through technology. Robinson, Cotabish, Wood, and O'Tuel (2014) observed gains in evaluation knowledge and skills, as did Moon (1996), through professional development during evaluation institutes. The DC-PET was able to achieve similar results using primarily an online application. Videos, links, and embedded resources provided the majority of the content and not a presenter at the front of the room.

Mixed-Methods Results

The first mixed-methods question stated, *“How did the qualitative results explain or expand on the quasi-experimental outcomes?”* The cross-case analysis was used to answer this question. Qualitative data collected during the focus groups, surveys,

and status checks supported the quantitative results already discussed. Many individuals remarked that they developed a newfound respect for evaluation and learned numerous skills from participating in the DC-PET process. In addition, the mean rating for evaluation team members' knowledge of program evaluation increased from 1.46 (0.61) on the pre-survey to 2.19 (0.57) on the post-survey using a four-point scale. This indicated a jump from novice to proficient, confirming the statistically significant results showing the frequency with which team members thought like an evaluator increased.

Fetterman and Wandersman (2005) described the dual benefits of participating in empowerment evaluation as being (a) gains in knowledge by the individuals who participate in the evaluation, and (b) gains in organizational knowledge leading to program improvement. The evidence for growth in personal gains described above and outlined in each case study is strong. The evidence for growth in organizational learning leading to program improvement is promising, but only time will tell. The development of SMART goals and the creation of a plan for communicating the results indicate an intention to make changes in the future, but does not necessarily mean the goals will be enacted. It was promising, however, to see four of the teams take immediate action to implement their SMART goals during the writing of this dissertation.

The second mixed-methods question stated, *What modifications should be made to the DC-PET in the future based on both the quantitative results and the findings generated from the qualitative data?* The cross-case analysis was used to answer this question. Three key modifications emerged from the data. The first change involves the creation of a new, more detailed introduction to the DC-PET. Second, electronic data entry capabilities should be incorporated into the online application. Third, additional

management tools for the leaders of evaluation teams using the DC-PET need to be developed.

One of the goals of the original Arkansas Evaluation Initiative (Robinson, Cotabish, Wood, & Biggers, 2009) was to collect evaluation exemplars. Now that the DC-PET has been piloted, I will be able to select high-quality examples of program descriptions, evaluation questions, communication plans, and final reports with the purpose of providing support and creating management tools to make the process even more comprehensible.

After reflecting on her study using the Purdue Three-Stage Model to provide professional development on evaluation to teachers and administrators, Moon (1996) wrote, “The most important resources for self-evaluations seemed to be knowledge about program evaluation, time, and computerization” (p. 126). Once the online application I created for this study has been modified with the suggestions made by the participants, a giant step will have been taken towards meeting the need for computerization identified by Moon almost 20 years ago.

Implications

There are three major implications of this study. The first implication is that, despite external evaluations being the gold standard, it is possible to conduct meaningful evaluations using internal evaluation methods as long as assistance is available from a knowledgeable individual as described in empowerment evaluation literature. This study provided evidence to show that engaging in empowerment evaluation can result in: (a) the participants in the evaluation gaining new knowledge and skills; and (b) the organization as a whole building long-term evaluation capacity. Empowerment

evaluation offers an inexpensive alternative to costly external evaluators and can be more critical and penetrating due to the attachment of the participants to the program (Fetterman & Wandersman, 2007).

The second implication of this study is that it is possible to build evaluation capacity primarily through a technology tool and not through intensive, on-site professional development of the typical sort. Information sharing has moved largely online due to ease of access, time savings, and usability. The DC-PET is currently the only evaluation tool designed for gifted educators available through an online application.

The third implication of this study is that the Kaplan Depth and Complexity Model can be used effectively to evaluate gifted education programs. The Kaplan Model is familiar to many in the field of gifted education and therefore lends to the comprehension of the new program evaluation knowledge users are asked to acquire. Teachers and administrators of gifted programs are hungry for user-friendly evaluation tools and truly desire to make decisions based on data instead of intuition alone. The Kaplan Model was designed to help individuals think like an expert in a discipline and can assist educators as they attempt to think and behave more like professional evaluators.

Limitations

The first limitation to this study is the sample size. A large number of participants is always desirable when analyzing quantitative data in particular. The sample size used to answer the quantitative question in this study was statistically adequate, but a larger sample would have been desirable.

The second limitation is that one-third of the participants had no experience evaluating a program in the past. This meant that when they were asked to compare the DC-PET to other methods of evaluation, they had little knowledge from which to draw. The fact that one-third of the participants were able to evaluate a program for the first time is positive evidence for the DC-PET's value, but did not help in answering qualitative research question one.

A third limitation to this study is that the comparison group consisted of only teachers and administrators of the gifted. The treatment group also included parents, other school staff, and students.

A fourth limitation is the attrition of evaluation team members. Attendance was not kept at each evaluation team meeting and I was not able to capture the thoughts and opinions of these individuals through the focus group or post-survey.

Opportunities for Future Research

The next logical steps for studying the effectiveness of the DC-PET is to make the modifications suggested by the participants and re-pilot the tool on a larger scale. It would be important to assess the availability of individuals to serve as empowerment evaluation coaches when I am removed from the picture. A job description would need to be created that would allow an individual to self-assess their readiness to fulfill that needed role.

A second important study I would like to conduct would be to follow-up with the original nine evaluation teams a year in the future. This study could examine (a) the accuracy of the force-field analysis conducted as a part of step 11, (b) the impact of the evaluation overall, and (c) whether or not the evaluation members maintained the

increased level of evaluative thinking they attained at the conclusion of this study. In essence, the study would examine if true change resulted from using the DC-PET as described by Lewin's (1951) change model (i.e. *unfreezing* the processes, *transitioning* to new ways of thinking and operating, and *refreezing* to ensure fidelity).

A third study I would like to conduct relates to the closeness in the relationship between evaluative thinking and critical thinking. It would be interesting to see if adding daily required critical thinking exercises to the DC-PET in addition to weekly group discussions of scenarios requiring evaluative thinking would further increase the degree to which participants use evaluative thinking as measured by the Evaluative Thinking Inventory. It would also be important to measure how this affects the quality of the evaluations that result from using the DC-PET.

A fourth study I would like to conduct would be to develop and pilot new tools for assessing the quality of an evaluation conducted using the DC-PET. The tool would incorporate the quality indicators discussed by Callahan (2006), the program evaluation standards developed by the Joint Committee on Standards for Educational Evaluation, and the evaluation standards included in the National Association for Gifted Children standards. In addition, more general questions will be included such as (a) *Did the evaluation team answer the evaluation questions they designed?* and (b) *Did the evaluation change the perspectives of the evaluators?* Having a tool like the one described would allow individuals to determine if the DC-PET assists users in overcoming the inertia of the constructivist principle, one of the appreciative inquiry principles described by Fetterman & Wandersman (2005), or inhibits an individual from seeing beyond the reality they have created for themselves.

Conclusion

Fetterman and Wandersman (2007) unequivocally stated, “empowerment evaluation is not a panacea” (p. 5). No evaluation approach is perfect for all individuals and in all circumstances. I also readily recognize that the Depth and Complexity Program Evaluation Tool (DC-PET) is not the holy grail of gifted program evaluation. However, I can confidently state as a result of this study that the DC-PET should be considered as a viable option and strong candidate for conducting a gifted program evaluation in the future. It is my hope that the DC-PET will help to fill the gaps in our understanding regarding the effectiveness of many common gifted program models (Van Tassel-Baska, 2006) one school at a time and call attention to the importance of program evaluation to the field as a whole.

REFERENCES

REFERENCES

- Ali, A . (2001). Issues involved in the evaluation of gifted programs. *Gifted Education International*, 16, 79-91.
- Aronson, E. & Patnoe, S. (2011). *Cooperation in the classroom: The jigsaw method* (3rd ed.). London: Pinter & Martin Ltd.
- Austin, M. & Harkins, D. (2008). Assessing change: Can organizational learning work for schools? *The Learning Organization*, 15(2), 105-125.
- Avery, L., & Van Tassel-Baska, J. (1997). Making evaluation work: One school district's experience. *Gifted Child Quarterly*, 41, 124-131.
- Baker, A., Bruner, B., Sabo, K. & Cook, A. (2006). *Evaluation capacity and evaluative thinking in organizations*. Retrieved from:
http://www.evaluativethinking.org/docs/EvalCap_EvalThink.pdf
- Barnard-Brak, L. Johnson, S., & Pond, A. (2009, August). *The incidence of potentially gifted students within a special education population*. Paper presented at the Biennial World Conference on Gifted and Talented Children, Vancouver, Canada.
- Baum, S. (2009). Talent centered model for twice exceptional students. In J. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press, Inc.

- Bell, L. (1986). Evaluation of the implementation of innovation in a gifted / enrichment program. *Gifted Education International*, 4, 120-127.
- Borland, J. (2009). Gifted education without gifted programs or gifted students: An anti-model. In J. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press, Inc.
- Brody, L. E., & Mills, C. J. (1997). Gifted children with learning disabilities: a review of the issues. *Journal of Learning Disabilities*, 30, 282-286.
- Brown, J. (2008). Statistics corner: Questions and answers about language testing statistics / Effect size and eta squared. Shiken: JALT Testing and Evaluation SIG Newsletter. 12(2), 38-43.
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*. Advance online publication. doi: 10.1177/1098214015581706.
- Buckley, J. & Archibald, T.(2011).*The Evaluative Thinking Inventory*. Cornell University. Retrieved from:
https://cornell.qualtrics.com/jfe/form/SV_eLKEC3fFDNuU8A
- Callahan, C. (1983). Issues in evaluating programs for the gifted. *Gifted Child Quarterly*, 27, 3-7.
- Callahan, C. (1986). Asking the right questions: The central issue in evaluating programs for the gifted and talented. *Gifted Child Quarterly*, 30, 38-42.

- Callahan, C. (2004). Program evaluation in gifted education (Vol.11). In S. M. Reis (Ed.) *Essential Readings in Gifted Education*. Thousand Oaks: CA: Corwin.
- Callahan, C. (2005). Developing a plan for evaluating a program in gifted education. In Purcell & Eckert (Eds.) *Designing Services and Programs for High-Ability Learners*. Thousand Oaks, CA: Corwin Press.
- Callahan, C. & Caldwell, M. (1997). *A practitioner's guide to evaluating programs for the gifted*. Washington, D.C.: National Association for Gifted Children.
- Callahan, C. & Hertberg-Davis. (2013). *Fundamentals of gifted education: Considering multiple perspectives*. New York: Taylor and Francis.
- Callahan, C., Moon, T., & Oh, S. (2014). National surveys of gifted programs: Executive Summary. *National Center on the Gifted and Talented*. Retrieved from:
<http://www.nagc.org/sites/default/files/key%20reports/2014%20Survey%20of%20GT%20programs%20Exec%20Summ.pdf>
- Carter, K. & Hamilton, W. (1985). Formative evaluation of gifted programs: A process and model. *Gifted Child Quarterly*, 29, 5-11.
- Carter, K. (1986). Evaluation design: Issues confronting evaluators of gifted programs. *Gifted Child Quarterly*, 30, 88-92.
- Carter, K. (1992). A model for evaluating programs for the gifted under non-experimental conditions. *Journal for the Education of the Gifted*, 15, 266-283.
- Cawsey, T., Deszca, G., & Ingols, C. (2012). *Organizational change: An action-oriented toolkit*. Thousand Oaks, CA: SAGE Publications.
- Clark, I. (2008). Assessment is for learning: Formative assessment and positive learning interactions. *Florida Journal of Educational Administration and Policy*, 2, 1-16.

- Clinkenbeard, P. (1996). Research on motivation and the gifted: Implications for identification, programming, and evaluation, *Gifted Child Quarterly*, 40, 220-221.
- Cooperrider, D. (1986). Appreciative inquiry: Toward a methodology for understanding and enhancing organizational innovation (Doctoral Dissertation, Case Western Reserve). UMI Number: 8611485.
- Cooperrider, D. & Whitney, D. (2005). *Appreciative inquiry: A positive revolution in change*. San Francisco: Berrett-Koehler Publishers, Inc.
- Cooperrider, L., Whitney, D., & Stavros, J. (2008). *AI handbook for leaders of change*. Brunswick, OH: Crown.
- Corbin, J. & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oak, CA: Sage Publications.
- Cotabish, A. & Robinson, A. (2012). The effects of peer coaching on the evaluation knowledge, skills, and concerns of gifted program administrators. *Gifted Child Quarterly*, 58, 160-170.
- Council of State Directors of Programs for the Gifted. (2013). State of the States in Gifted Education. Washington DC: National Association for Gifted Children (NAGC).
- Cousins, J. & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, No. 80, 5-23.
- Covey, S. (1990). *Seven habits of highly effective people*. New York: Simon & Schuster.
- Creswell, J. (2007). *Qualitative Inquiry and Research Design: Choosing among five approaches*. Thousand Oaks, CA: Sage Publications, Inc.

- Creswell, J. & Clark, P. (2007). *Designing and conducting mixed-methods research*. Thousand Oaks, CA: Sage Publications, Inc.
- Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Dai, D., Swanson, J. and Cheng, H.(2011). State of research on giftedness and gifted education: A survey of empirical studies published during 1998-2010. *Gifted Child Quarterly*, 55, 126-138.
- Davidson, E., Howe, M., & Scriven, M. (2004). Evaluative thinking for grantees. In M. Braverman, N. Constantine, & J.K. Slater (Eds.), *Foundations and evaluation: Contexts and practices for effective philanthropy* (pp. 259-280). San Francisco, CA: Jossey-Bass.
- Dickerson, M. (2011). Reculturing schools for greater impact: Using appreciative inquiry as a non-coercive change process. *International Journal of Business and Management*, 6(8),66-74.
- Dodds, K. (2010). Effects of the prompts of depth and complexity on gifted and non-gifted students (Doctoral Dissertation, University of Southern California). UMI Number: 3418029.
- Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*, 70, 35–36.
- Doveston, M. & Keenaghan, M.(2006). Improving classroom dynamics to support students' learning and social inclusion: A collaborative approach. *Schools for Learning*, 21, 5-11.

- Eddles-Hirsch, K. (2012). Insiders or outsiders: The role of social context in the peer relations of gifted students. *Roeper Review*, 34, 53-62.
- Fetterman, D. & Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York: Guilford Press.
- Fetterman, D. & Wandersman, A. (2007). Empowerment evaluation: Yesterday, today, and tomorrow. *American Journal of Evaluation*, 28, 179-198.
- Fleischer, D., & Christie, C. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, 30, 158-175.
- Ford, D. (1998). The underrepresentation of minority students in gifted education. *Journal of Special Education*, 32, 4-14.
- Ford, D., Tarek, C., & Gilman, W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children*, 74, 289-306.
- Gentry, M. & Gable, R. (2000). *My Class Activities*. Mansfield Center, CT: Creative Learning Press.
- Harmon, R., Fontaine, D. Plews-Ogan, M., & Williams, A. (2012). Achieving transformational change: Using appreciative inquiry for strategic planning in a school of nursing. *Journal of Professional Nursing*, 28, 119-124.
- Higgs, M. & Rowland, D. (2005). All changes great and small: Exploring approaches to change and its leadership. *Journal of Change Management*, 5, 121-151.
- House, E., & Lapan, S. (1994). Evaluation of programs for disadvantaged gifted students. *Journal for the Education of the Gifted*, 17, 441-466.

Hunsaker, S. (2000). Documenting gifted program results for key decision-makers.

Roeper Review, 23, 80-82.

Hunsaker, S., & Callahan, C. (1993). Evaluation of gifted programs: Current practices.

Journal for the Education of the Gifted, 16, 190-200.

Jacobs, H. & Brandt, D. (2012). Opening up a 'third space' to innovate curricula. *Journal of Social Sciences*, 8, 420-428.

Johnsen, S. (2000). What the research says about accountability and program evaluation.

Tempo, 23-30.

Johnsen, S. (2011). *Identifying Gifted Students: A Practical Guide*. Waco, TX: Prufrock Press Inc.

Johnsen, S. (2014). Gifted programming standards. In J. Plucker & C. Callahan (Eds.)

Critical issues and practices in gifted education: What the research says (2nd ed.) (pp. 281-296). Waco, TX: Prufrock Press.

Jones, B. (2011) The Texas Academy of Mathematics and Science: a 20-year perspective.

Journal for the Education of the Gifted, 34, 513-543.

Kaplan, S. (2009). The grid: A model to construct differentiated curriculum for the gifted.

In J. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented (2nd ed.)* (pp. 235-252). Mansfield Center, CT: Creative Learning Press, Inc.

Kotter, J. (2007). Leading change: Why transformation efforts fail. *Harvard Business*

Review, 85, 96-103.

- Labin, S., Duffy, J., Meyers, D., Wandersman, A., & Lesesne, C. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33, pp. 307-338.
- Lauer, J. L. (2010). *Experts' perspectives on the application and relevancy of depth and complexity to academic disciplines of study*. (Order No. 3418261, University of Southern California). *ProQuest Dissertations and Theses*, 141. Retrieved from <http://search.proquest.com/docview/751224867?accountid=13360>. (751224867).
- Lewin, K. (1951). *Field theory in social science*. New York: Harper and Row.
- Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109.
- Marland, S. (1972). Education of the gifted and talented: Report to the Congress of the United States Washington DC.
- McBee, M. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Journal of Secondary Gifted Education*, 17, 103-111.
- McCoach, B., Gable, R., & Madura, J. (2013). *Instrument development in the affective domain: School and corporate applications*. New York: Springer Science + Business Media.
- McIntosh, J. (2014). *The depth and complexity program evaluation tool (DC-PET)*. Unpublished manuscript.

- McIntosh, J. (2015). Gifted Program Evaluation: A Handbook for Administrators and Coordinators. Speirs Neumeister, K. & Burney, V. (2012), *Roeper Review*, 37:2, 111-112.
- Mertens, D. & Hesse-Biber, S. (2013). Mixed methods and credibility of evidence in evaluation. *New Directions for Evaluation*, 138, 5-13.
- Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, 27, 296-319.
- Mintzberg, H. & Westley, F. (1992). Cycles of organizational change. *Journal of Strategy Management*, 13, 39-59.
- Moon, S. (1996). Using the Purdue three-stage model to facilitate local program evaluations. *Gifted Child Quarterly*, 40, 121-128.
- NAGC. (2010). Pre-K-Grade 12 gifted programming standards. Retrieved from: http://www.nagc.org/uploadedFiles/Information_and_Resources/Gifted_Program_Standards/K-12%20programming%20standards.pdf
- Neuman, S. & McCormick, S. (1995). *Single-subject experimental research: Applications for literacy*. Newark, DE: International Reading Association.
- Norum, K.(2001).Appreciative design. *Systems Research and Behavioral Science*, 18, 323-333.
- Orlikowski, W. (1996). Improving organizationoal tranformation overtime: A situated change perspective. *Information Systems Research.*, 7, 63-92.
- O’Sullivan, R. (2004). *Practicing evaluation: A collaborative approach*. Thousand Oaks, CA: Sage.

- Owen, J. (2007). *Program evaluation: Forms and approaches* (3rd ed.). New York: The Guilford Press.
- Passow, A. & Frasier, M. (1996). Toward improving identification of talent potential among minority and disadvantaged students. *Roeper Review*, 18, 198-202.
- Paul, K. A. (2010). A national study of state policy for fostering local gifted program evaluation: Content analysis and recommendations for policy development (Doctoral Dissertation, University of Connecticut). UMI Number: 3464370.
- Peterson, J., Assouline, S, & Jen, E. (2015) Responding to concerns related to the social emotional development of gifted adolescents. In F. Dixon & S. Moon (Eds) *The handbook of secondary gifted education* (pp. 65-90). Waco, TX: Prufrock Press.
- Peterson, J. S., & Moon, S. M. (2008). *Counseling the gifted*. In S. Pfeiffer (Ed.), *Handbook of Giftedness in Children: Psycho-Educational Theory, Research, and Best Practices* (pp. 225-248). New York, NY: Springer.
- Pfeiffer, S. I., & Stocking, V. B. (2000). Vulnerabilities of academically gifted students. *Special Services in the Schools*, 16, 83–93.
- Plucker, J. & Callahan, C. (2013). *Critical issues and practices in gifted education, 2E: What the research says*. Waco, TX: Prufrock Press.
- Reichardt, C. (2011). Evaluating methods for estimating program effects. *American Journal of Evaluation*, 32, 246-272.
- Renzulli, J. (1975). *A guidebook for evaluating programs for the gifted and talented*. Camarillo, CA: Ventura County Supt of Schools.

- Renzulli, J., Gubbins, J., McMillen, K., Eckert, R., & Little, C. (2009). *Systems and models for developing programs for the gifted and talented*, 2nd edition. Waco, TX: Prufrock Press.
- Robinson, A., Cotabish, A., Wood, B., & Biggers, A. (2009). The Arkansas evaluation initiative in gifted education. In J. Renzulli (Ed.), *Systems and Models for Developing Programs for the Gifted and Talented*. Mansfield, CT: Creative Learning Press, Inc.
- Robinson, A. Cotabish, A., Wood, B., & O'Tuel, F. (2014). The effects of a statewide evaluation initiative in gifted education on practitioner knowledge, concerns, and program documentation. *Journal of Advanced Academics*, 25, 349-383.
- Scheirer, M. (2012). Planning evaluation through the program life cycle. *American Journal of Evaluation*, 33, 263-294.
- Schon, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Schwandt, T. (2008). Educating for intelligent belief in evaluation. *American Journal of Evaluation*, 29, 139-150.
- Seeley, K. (1986). Evaluation for defensible programs for the gifted. In J. Maker (Ed.), *Critical Issues in Gifted Education: Defensible Programs for the Gifted*. New York: Aspen Publishers.
- Silverman, L. (2013). Asynchronous development: Theoretical basis and current applications. In C.S. Neville, M. M. Piechowski, and S. S. Tolan (Eds.) *Off the charts: Asynchrony and the gifted child* (pp. 18-47). Unionville, NY: Royal Fireworks Press.

- Simonsen, B., Little, C., & Fairbanks, S. (2010). Task difficulty, attention, and student behavior: A study of student behavioral response during differentiated tasks. *Journal for the Education of the Gifted*, 34, 245-260.
- Slavin, R. (2008). Perspectives on evidence-based research in education: What works? Issues in Synthesizing educational program evaluations. *Educational Researcher*, 37, 5-14.
- Speirs Neumeister, K. & Burney, V. (2012). *Gifted Program Evaluation: A Handbook for Administrators and Coordinators*. Waco, TX: Prufrock Press.
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation*, 89, 1-99.
- Swanson, J. & Lord, W. (2013). Harnessing and guiding the power of policy: Examples from one state's experiences. *Journal for the Education of the Gifted*, 36, 198-219.
- Tomlinson, C., Bland, L., & Moon, T. (1993). Evaluation utilization: A review of the literature with implications for gifted education *Journal for the Education of the Gifted*, 16, 171-189.
- Tomlinson, C., & Callahan, C. (1993). Planning effective evaluations for programs for the gifted. *Roeper Review*, 17, 46-51.
- Tomlinson, C. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: ASCD.

- Tomlinson, C., Brighton, C., Hertberg, H., Callahan, C., Moon, T., Brimijoin, K., Conover, L., & Reynolds, T. (2003). Differentiation instruction in response to students readiness, interest, and learning profiles in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, 27, 119-145.
- Tschannen-Moran, M. & Tschannen-Moran, B.(2011).Taking a strengths-based focus improves school climate. *Journal of School Leadership*, 21, 422-448.
- U.S. Department of Education. (1993). *National excellence: A case for developing America's Talent*. Washington, DC: US Government Printing Office.
- VanTassel-Baska, J. & Feng, A. (2004). *Designing and utilizing evaluation for gifted program improvement*. Waco, TX: Prufrock Press.
- Van Tassel-Baska, J. (2006). A content analysis of evaluation findings across 20 gifted programs: A clarion call for enhanced gifted program development. *Gifted Child Quarterly*, 50, 199-215.
- VanTassel-Baska, J., McIntosh, J., & Kearney, K. (2015). Secondary affective curriculum and instruction for gifted learners. In F. Dixon & S. Moon (Eds) *The handbook of secondary gifted education* (pp. 509-539). Waco, TX: Prufrock Press.
- Vygotsky, L. (1986). *Thought and language* (A. Kozuin, Trans.). Cambridge, MA: MIT Press.
- W. K. Kellogg Foundation. (1998). *Using logic models to bring together planning, evaluation, and action: Logic development guide*. Battle Creek, MI: W.K. Kellogg Foundation.

- Walsh, R. & Kemp, C. (2013). Evaluating interventions for young gifted children using single-subject methodology: A preliminary study. *Gifted Child Quarterly*, 57, 110-120.
- Wandersman, A., Snell-Johns, J., Lentz, L., Fetterman, D. M., Keener, D. C., Livet, M., et al. (2005). The principles of empowerment evaluation. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice* (pp. 27-41). New York: Guilford.
- Warne, R. (2014). Using above-level testing to track growth in academic achievement in gifted students. *Gifted Child Quarterly*, 58, 3-23.
- Wood, S. M. (2012). Rivers' confluence: A qualitative investigation into gifted educators' experiences with collaboration with school counselors. *Roeper Review*, 34, 261-274.
- Worrell, F. (2015). Being gifted and adolescent: Issues and needs of diverse students. In Dixon & Moon (Eds.) *The handbook of secondary gifted education*, 2nd edition. Waco, TX: Prufrock Press.
- Yarbrough, D., Shulha, L., Hopson, R., & Caruthers, F. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. Thousand Oaks, CA: Sage Publications, Inc.
- Yin, R. (2014). *Case study research: Design and methods*. Thousand Oaks, CA: SAGE Publications.
- Yoon, S. & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly*, 53, 121-136.

APPENDICES

Appendix A

Depth and Complexity Program Evaluation Tool (DC-PET)

WORKING DRAFT- Jason McIntosh


The Depth and Complexity
Program Evaluation Tool (DC-PET)

A Workbook for Using the Kaplan Depth and Complexity Model to
Conduct an Internal Gifted Program Evaluation

1. To the best of your ability, describe your gifted program as it is today.

Include in your description the identification process, the types of programs offered, the curriculum used, and any other important details.

Origins

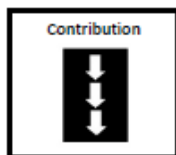


When your program is working at its best, what does this look like? Think about a time when a gifted student, parent, or teacher shared a positive story with you about their experience with the gifted program. Write down a short narrative in the box below and share it with someone else.

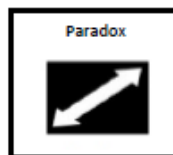
Note. Do not use all or part of the DC-PET without permission from the author.

2. What is the overall purpose of the evaluation?

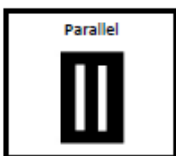
Four common purposes are listed below. Circle the icon that applies to this evaluation or write a different purpose on the line provided.



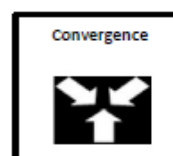
- To measure the impact or effectiveness of a program



- To uncover any weaknesses or contradictory practices



- To compare a program to the NAGC Standards or to assess compliance with state laws

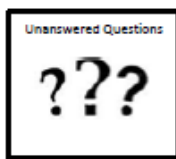


- To determine the factors that have lead to the current perceptions of a program

OTHER:

3. What unanswered questions do you have about the program you are evaluating?

Brainstorm a list below:



Program Evaluation
Standards: U5, P1, P4

NOTE: The standards listed below each icon pertain to the Joint Committee on Standards for Educational Evaluation Program Evaluation Standards.

Examples:

1. Are the identification methods aligned with the services that are provided?

2 Is the curriculum rigorous enough?

Note. Do not use all or part of the DC-PET without permission from the author.

4. Have you considered how your gifted program addresses the four equity and excellence issues listed below?



Program Evaluation
Standards: U4, U6

Stop and reflect for a moment on the four equity and excellence issues shown in the box below. A book or article citation relating to each topic has been provided as a starting point for discussion. Select one to examine more closely throughout the evaluation. Prompts will be provided periodically to focus your attention back on this issue.

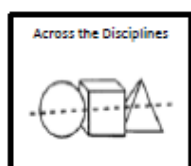
✓	Issue	Resource
	Underrepresented populations	Ford, D. (1998). <i>The underrepresentation of minority students in gifted education</i> . <i>Journal of Special Education</i> , (32)1, pp. 4-14.
	Twice-exceptional learners	Baum, S. (2009). <i>Talent centered model for twice exceptional students</i> . In J. Renzulli (Ed.) <i>Systems and models for developing programs for the gifted and talented</i> (pp. 17-48). Mansfield Center, CT: Creative Learning Press.
	Affective needs of students	Peterson, J. S., & Moon, S. M. (2008). <i>Counseling the gifted</i> . In S. Pfeiffer (Ed.), <i>Handbook of Giftedness in Children: Psycho-Educational Theory, Research, and Best Practices</i> (pp. 225-248). New York, NY: Springer.
	Defensible differentiation	Borland, J. (2009). <i>Gifted education without gifted programs for gifted students: An anti-model</i> . In J. Renzulli (Ed.) <i>Systems and models for developing programs for the gifted and talented</i> (pp. 105-118). Mansfield Center, CT: Creative Learning Press, Inc.

5. Which of your unanswered questions will become the focus of the evaluation?

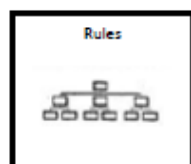
Use the matrix below to help you narrow down your choices.

Callahan (1986) describes the characteristics of a good evaluation question in the following three ways: →	Relevant= address the functioning of the program, its components, goals, or structure	Useful= provide data that some audience can actually use to make decisions.	Important= yield data helpful in making decisions which will have an impact.
List of top 5 possible questions:	Is the question relevant?	Is the question useful?	Is the question important?
1.	Yes or No	Yes or No	Yes or No
2.	Yes or No	Yes or No	Yes or No
3.	Yes or No	Yes or No	Yes or No
4.	Yes or No	Yes or No	Yes or No
5.	Yes or No	Yes or No	Yes or No

Note. Do not use all or part of the DC-PET without permission from the author.



NOTE: If one of your chosen evaluation questions relates to the curriculum gifted students in your district use, make sure to consider the "Across the Disciplines" icon. Is the curriculum interdisciplinary, rigorous, concept-based, engaging, and student centered? Have the common core standards been addressed?



Choose two of the most relevant, useful, and important questions from the box above and rewrite them in boxes one and two below. Rewrite the equity and excellence issue you selected earlier in box three as well. These will become your evaluation questions.

Final Evaluation Questions	
1.	(UNANSWERED QUESTION)
2.	(UNANSWERED QUESTION)
3.	(EQUITY AND EXCELLENCE ISSUE)

6. What data collection procedures will you use?



Collect information pertaining to your evaluation questions from all stakeholder groups including but not limited to parents, students, teachers, and administrators. This will help you triangulate the data to ensure it is accurate and representative.

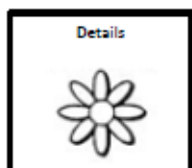
Program Evaluation
Standards: U2, U5, F3, P1, P3

Place a checkmark in the box next to the data collection method you will use for each stakeholder group.

	Teachers	Parents	Students	Administrators
Survey				
Interview				
Focus Group				
Observation				
Needs Assessment				
Document Review				
Other:				

Note. Do not use all or part of the DC-PET without permission from the author.

7. When will the evaluation take place and who will be responsible for each step?



Details

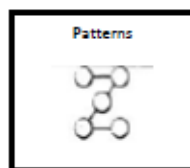
Use the table below to plan when each step of the evaluation will take place and by whom. Phase one consists of how the data will be collected. Phase two, outlined in step eight, will help you analyze the data you will collect.

Program Evaluation
Standards: F4, A5, A7, U7

Phase One					
Data Collection Strategy	Date	Person Responsible	Data Collection Strategy	Date	Person Responsible

8. What patterns and trends do you see? (*Phase Two*)

Data can be analyzed in several ways. One useful method is to look for patterns and trends that emerge. Kaplan (2009) describes patterns as reoccurring events and trends as influences or forces that shape ideas. Use the box below to record any patterns or trends that you discover.



Patterns

Program Evaluation
Standards: U6, A7

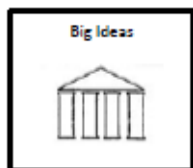


Trends

Note. Do not use all or part of the DC-PET without permission from the author.

Before proceeding, download and print the current National Association for the Gifted (NAGC) standards from the following website - www.nagc.org/ProgrammingStandards.aspx. Use a highlighter to identify the NAGC standards that pertain to your evaluation questions.

9. What conclusions can be made from the patterns and trends discovered in step eight?



Program Evaluation
Standards: P5, A1, A2,
A3, A7

Take time now to objectively compare and contrast the patterns and trends you identified within your gifted program to the NAGC standards highlighted earlier. Use the box below to generate a list of the strengths and weaknesses of your gifted program.

<i>Strengths of the Program</i>	<i>Areas for Improvement</i>

10. What are the next steps for the program based on the strengths and weaknesses identified above?



Program Evaluation
Standards: U4, U6, U8, A7

Use the table on the next page to brainstorm a list of possible recommendations for change. Make sure to address the equity and excellence issue identified in step four. Rank order your ideas from highest priority to lowest priority.

Note. Do not use all or part of the DC-PET without permission from the author.

<i>Possible Recommendations</i>	<i>Rank Order</i>

Choose two or three of your highest priority recommendations and create an action plan for making them happen. This can be accomplished through the creation of SMART goals. SMART is an acronym for *Specific, Measurable, Attainable, Relevant, and Timely*. For more information on writing SMART goals, click the following link- http://en.wikipedia.org/wiki/SMART_criteria

Example SMART Goal: *All gifted education teachers in the district will meet with each grade-level team at their school during the month of October to explain the new gifted and talented screening process.*

SMART Goal #1	
SMART Goal #2	
SMART Goal #3	

Before proceeding, ask yourself if each goal is indeed SMART! Revisit the website to make sure.

Note. Do not use all or part of the DC-PET without permission from the author.

11. How will you communicate your findings to the stakeholders in a clear, jargon free manner that is appropriate for each audience?



Program Evaluation
Standards: U7, P5, P6, A8

Communication plans are important for several reasons:

- Pre-change approval,
- Create need for change,
- Midstream and milestone change,
- Confirm and celebrate success

-(Cawsey, Descza, & Ingols, 2012)

In your opinion, what are the forces for and against change in your program?

Forces FOR Change	Forces AGAINST Change

Think about this as you decide what information each stakeholder needs to know and how you will present it. A few of the most common methods of presentation are:

-PowerPoint -Poster -Pamphlet -Flyer
-Podcast -Video -Website -Lecture

Use the box below to create a draft communications plan.

	Teachers	Parents	Students	Administrators
Important Information to Share				
Presentation Method/s				

FINAL NOTE: A formal written report is often required at the conclusion of an evaluation. Access to this report should be given to all stakeholders and generally includes the following:

-Title page -Table of contents -Introduction -Background
-Methodology -Results -Conclusion -Recommendations
-References -Appendices -Executive summary

Note. Do not use all or part of the DC-PET without permission from the author.

Appendix B.

Evaluative Thinking Inventory

Evaluative Thinking Inventory

ID #:

- Please read each of the statements below and check the appropriate box to indicate how often you do what is described by each statement.
- When applicable (depending on the question), consider how you think and act in both professional and personal settings.

	Very Frequently	Frequently	Occasion- ally	Rarely	Very Rarely	Never
1. I describe my thinking to others.						
2. I am eager to engage in evaluation.						
3. I suggest alternative explanations and hypotheses.						
4. I use models and/or other diagrams to clarify my thoughts.						
5. I take time to reflect about the way I do my work.						
6. I discuss evaluation strategies with my colleagues.						
7. I articulate the logical justification of my evaluation strategy.						
8. I consider alternative explanations for claims.						
9. I am wary of claims made by others without evidence to back them up.						
10. I seek evidence for claims and hypotheses.						
11. I am interested in understanding the logic behind things.						
12. I articulate the relationship between my evaluation work and my intended claim						
13. I reflect on assumptions and claims I make myself.						
14. I pose questions about assumptions and claims made by others.						
15. I willingly make changes to the way I do my work.						
16. I try to convince others that evaluation is important.						
17. I offer evidence for claims that I make.						
18. I use models and/or other diagrams to communicate my thinking to others.						
19. I believe evaluation is a valuable endeavor.						
20. I enjoy discussing evaluation strategies with colleagues.						

Appendix C.

American Evaluation Association's Program Evaluation Standards

Utility Standards

U1- Evaluator Credibility	Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
U2- Attention to Stakeholders	Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
U3- Negotiated Purposes	Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
U4- Explicit Values	Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
U5- Relevant Information	Evaluation information should serve the identified and emergent needs of stakeholders.
U6- Meaningful Processes and Products	Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.
U7- Timely and Appropriate Communicating and Reporting	Evaluations should attend to the continuing information needs of their multiple audiences.
U8- Concern for Consequences and Influence	Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

Feasibility Standards

F1- Project Management	Evaluations should use effective project management strategies.
F2- Practical Procedures	Evaluation procedures should be practical and responsive to the way the program operates.
F3- Contextual Viability	Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.
F4- Resource Use	Evaluations should use resources effectively and efficiently.

Propriety Standards

P1- Responsive and Inclusive Orientation	Evaluations should be responsive to stakeholders and their communities.
P2- Formal Agreements	Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.
P3- Human Rights and Respect	Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and others stakeholders.
P4- Clarity and Fairness	Evaluations should be understandable and fair in addressing stakeholder needs and purposes.
P5- Transparency and Disclosure	Evaluations should provide complete descriptions of findings, limitations, and conclusion to all stakeholders, unless doing so would violate legal and propriety obligations.

P6- Conflicts of Interest	Evaluations should openly and honestly identify and address real or perceived conflicts of interest that may compromise the evaluation.
P7- Fiscal Responsibility	Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

Accountability Standards

E1- Evaluation Documentation	Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.
E2- Internal Metaevaluation	Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.
E3- External Metaevaluations	Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external metaevaluations using these and other applicable standards.

Accuracy Standards

A1- Justified Conclusions and Decisions	Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.
A2- Valid Information	Evaluation information should serve the intended purposes and support valid interpretations.
A3- Reliable Information	Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.
A4- Explicit Program and Context Descriptions	Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.

A5- Information Management	Evaluations should employ systematic information collection, review, verification, and storage methods.
A6- Sound Designs and Analyses	Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.
A7- Explicit Evaluation Reasoning	Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.
A8- Communication and Reporting	Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

Appendix D.

Permission to use the 'Evaluative Thinking Inventory'



Department of Agricultural, Leadership, and
Community Education. (MC0343)
Litton Reaves, Room 214
175 W. Campus Drive
Blacksburg, Virginia 24061
540-231-6836 Fax: 540-231-3824
www.alce.vt.edu

June 17, 2015

Dear Jason McIntosh,

I am writing to officially confirm that Jane Buckley and I granted you permission to use the Evaluative Thinking Inventory in your dissertation study. As you know, we developed the scale based on theory and presented on it in a number of venues (e.g., Archibald & Buckley, 2011, 2012) and have subsequently developed our related work on promoting evaluative thinking (Buckley, Archibald, Hargraves, & Trochim, 2015), but had not used the scale with a large enough sample size to examine the measure's quality. As such, we are very happy that your dissertation study is making use of the scale and contributing to establishing it as a quality measure that others in the field can use in the future.

Sincerely,

Thomas Archibald
Assistant Professor

- Archibald, T., & Buckley, J. (2012, May). Evaluative Thinking: The 'Je Ne Sais Quoi' of Evaluation Capacity and Evaluation Quality, presented to the American Evaluation Association 'coffee break' webinar series.
- Archibald, T., Buckley, J., & Trochim, W. (2011, November). Evaluative Thinking: What is it? Why does it Matter? How Can We Measure It? Paper presented at the annual conference of the American Evaluation Association, Anaheim, CA.
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*. Advance online publication. doi:10.1177/1098214015581706.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution

Appendix E.

Master Checklist of Gifted Program Elements for Self-Assessment



Master Checklist of Gifted Program Elements for Self-Assessment

Program Design Items	No Evidence	Some Evidence	In Place	Comments
1. There is a written philosophy and/or mission statement related to gifted students.				
2. There is a written definition of which students the district considers to have what particular needs that require specialized services.				
3. There are written goals and objectives for these services.				
4. There is a written description of the services to be provided for the described students at each grade level and in each area served.				
5. Services provided align with how giftedness is defined.				
6. Gifted students are grouped together for instruction in their area(s) of talent.				
7. Services are constructed so that there is a continuum of services to meet the broad range of needs of individual gifted students.				
8. Policies are in place to allow early entrance, grade skipping, subject skipping, early credit, and early graduation according to individual student need.				
9. The roles of personnel at the district, the building and the classroom are clearly defined.				
10. A district-wide stakeholder group exists and meets on a regular basis to review the district services for gifted students.				

Identification Items	No Evidence	Some Evidence	In Place	Comments
11. The district uses a norm-referenced measure of ability in each of the areas for which program services are offered (i.e. math, language arts).				
12. The district uses a norm-referenced measure of achievement with adequate ceilings to assess achievement above grade level in each of the areas for which program services are offered.				
13. The district uses qualitative indicators of ability to perform in each of the areas for which program services are offered.				
14. The procedures ensure that all students have an opportunity to be nominated for screening by publicizing the process and receiving nominations from all stakeholder groups.				
15. Students are identified in all grade levels for which services are provided.				
16. The formal identification process is repeated at targeted grade levels including (but not limited to) kindergarten, 2 nd grade, prior to placement for middle school, and prior to placement in high school.				
17. The appeals process is publicized.				
18. The appeals process allows for students to take alternative ability, achievement, and/or qualitative measures at no cost to the family.				
19. The exit procedure includes period of intervention no less than one grading period to determine if student can be successful in the program with supports.				
Curriculum and Instruction Items	No Evidence	Some Evidence	In Place	Comments
20. There is a written curriculum in core subject areas and other areas served by the district that is specific to students identified as gifted K-12.				
21. Student learning goals are clear, and evidence of how the learning will be demonstrated is clearly stated.				

22.	The written curriculum has clear evidence of vertical articulation from grade to grade and K-12.				
23.	There is clear evidence of acceleration of curriculum in areas served.				
24.	There is clear evidence of enrichment of curriculum in areas served.				
25.	Instruction and learning experiences are clearly differentiated to focus on higher order thinking.				
26.	There is evidence of teaching of communication, collaboration, research, critical thinking, problem solving.				
27.	The pace of instruction is appropriate for gifted students.				
28.	There is evidence of student use of technology for creating content, learning content, and communicating content.				
29.	Assessments are aligned to curriculum goals.				
30.	Pre-assessment is used to determine individual instructional plans.				
31.	Post-assessment is used to demonstrate student growth and attainment of stated learning goals.				
Affective Needs Items		No Evidence	Some Evidence	In Place	Comments
32.	A written, differentiated, affective curriculum is available and used by teachers that addresses social and emotional needs of gifted students.				
33.	Affective curriculum teaches students about social and emotional characteristics as well as potential issues they may face.				
34.	Documentation of differentiated college guidance for gifted students is available (e.g. fieldtrips, independent study projects, speakers, or shadowing experiences pertaining to college exploration).				
35.	Documentation of differentiated career guidance for gifted students is available (e.g. fieldtrips, independent study projects, mentors, speakers,				

or shadowing experiences pertaining to college exploration).				
Professional Development Items	No Evidence	Some Evidence	In Place	Comments
36. Personnel working with gifted students are provided with opportunities for continuing professional development in the area of gifted education.				
37. Parents of gifted students are provided with opportunities for professional development about the characteristics and needs of this population.				
Program Evaluation Items	No Evidence	Some Evidence	In Place	Comments
38. The district uses multiple strategies to assess gifted student performance and growth.				
39. All components of the high ability program are periodically reviewed by individuals knowledgeable about gifted learners and who have competence in the evaluation process. The results are used for continuing program improvement.				
40. The evaluation report for all educational services involving gifted students includes both strengths and areas of challenge of the program and is accompanied by a plan with implications for improvement and renewal over time.				
41. The results of the program evaluation are presented to the local school board, the stakeholder group, and accessible to all constituencies of the program.				

Checklist is taken from Neumeister, K. S., & Burney, V. (2012). *Gifted program evaluation: A handbook for administrators & coordinators*. Waco, TX: Prufrock Press. Used with permission of Prufrock Press.

VITA

VITA

Jason McIntosh

mcintosh.jason1@gmail.com

EDUCATION

- 2015 Doctoral of Philosophy, Purdue University, West Lafayette, Indiana
Major Area: Educational Psychology, Gifted and Talented Education
Advisor: Dr. Marcia Gentry

- 2008 Miami University, Oxford, Ohio
Master of Arts in Teaching
Major Area: Biological Sciences

- 2000 Miami University, Oxford, Ohio
Master of Education
Major Area: Elementary Education

- 1996 Miami University, Oxford, Ohio
B.S., Education
Major Area: Elementary Education

TEACHING CREDENTIALS

- 2007 – 2016 Arizona Certified Teacher - K-8 Standard Education, K-12 Gifted Education

- 2003 – 2013 National Board Certified Teacher – Middle Childhood / Generalist

UNIVERSITY TEACHING EXPERIENCE

- Fall 2015 Teaching Assistant for the course “Creativity: Research and Development in Gifted Education”
Arizona State University, Tempe, Arizona
Overseeing Professor: Dr. Dina Brulles

- Summer 2015 Teaching Assistant for the course “The Gifted Learner: Foundations of Identification, Education, and Development”
Arizona State University, Tempe, Arizona
Overseeing Professor: Dr. Dina Brulles

- Fall 2013 Teaching Assistant for the course “Creating Learning Environments”
Purdue University, West Lafayette, Indiana
Overseeing Professor: Dr. Helen Patrick
- Fall 2013 Teaching Apprentice for the course “Curriculum and Program Development in Gifted Education”
Purdue University, West Lafayette, Indiana
Overseeing Professor: Dr. Kristina Ayers Paul
- 2011 - 2012 Teaching Assistant for the course “Creating Learning Environments”
Purdue University, West Lafayette, Indiana
Overseeing Professor: Dr. Eric Mann
- 2007 – 2010 Instructor: Gifted Endorsement Courses (see below)
Contracted by Chapman University and Educational Capstones Consulting
- Introduction to Gifted Education (Fall of 2007, 2008, and 2009)
 - Teaching the Gifted Learner (Fall of 2007, 2008, and 2009)
 - Effective Program Models in Gifted Education (Spring of 2008-2010)
 - Dual Labels in Gifted Education (Spring of 2008- 2010)
 - Social and Emotional Needs of Gifted Learners (Spring of 2008-2010)

K-12 TEACHING EXPERIENCE

- 2007 – 2011 Collaborative Peer Teacher for the Gifted
Avondale Elementary School District, Avondale, Arizona
- Coordinated all aspects of the gifted program, which included providing formal and informal professional development to 85 gifted cluster teachers, conducting a yearly program evaluation, holding parent meetings, organizing after school enrichment activities, and overseeing all gifted testing and scoring.
 - Served on the Avondale School Improvement Plan Walkthrough Team and various other committees including a hiring committees, textbook adoption committee, and the Kids at Hope Implementation Team
- 2005 – 2007 1st-5th Grade Gifted and Talented Pull-out Teacher
Lake Windward Elementary School, Fulton County, Georgia
- Provided enrichment and acceleration to gifted students in the areas of science and social studies. Administered and scored four different gifted identification instruments throughout the year.
- 2004 – 2005 3rd Grade Classroom Teacher
Antilles School, St. Thomas, U.S. Virgin Islands
- Taught 17 multi-national students in a self-contained classroom.

- 2001 – 2004 Gifted Intervention Specialist
Franklin City Schools, Franklin, Ohio
- Provided enrichment and acceleration to all 4th-6th grade gifted students in the district. Collaborated and consulted with regular education teachers in regards to differentiation and gifted education.
- 1996 – 2001 3rd and 4th Grade Classroom Teacher
Hunter Elementary School, Franklin, Ohio
- Taught a self-contained classroom in two year loops.

RELATED EXPERIENCE

- 2011 - 2015 Research Assistant, Gifted Education Resource Institute, Purdue West Lafayette, IN
- Coordinated Summer Residential program, gathered data for the HOPE+ Grant, contributed to the creation of online modules for professional development
- 1999 – 2003 Warren County Summer Nature Program Creator and Facilitator
Franklin, OH
- Proposed, created, and obtained funding and approval for a summer nature program by the Warren County Board of Trustees.

SERVICE

- 2015 NAGC National Convention Program Committee
- Select presenters and design schedule for the national conference
- 2014 - 2015 Board Member, Arizona Association for Gifted and Talented
- State Conference Chair, Executive Committee voting member, financial officer
- 2014 President, Graduate Organization in Educational Studies
Purdue University
- Lead group meetings, serve as university representative for GOEDS
- 2013 Graduate Student Representative, Research on Giftedness, Creativity, and Talent SIG
American Educational Research Association (AERA)
- Organize SIG social and provide support to board
- 2013 Vice-President, Graduate Organization in Educational Studies
Purdue University
- Organize and plan events throughout the year

- 2012 - 2015 Membership Chair, Professional Development Network
National Association for Gifted Children (NAGC)
-Contact new and renewing members, write column for network newsletter
- 2012 Membership Committee, Research and Evaluation Network
National Association for Gifted Children (NAGC)
-Contribute to the “Did You Know?” series and plan webinars
- 2012 Division C Graduate Student Committee Campus Liaison for Purdue University
American Educational Research Association (AERA)
-Disseminate information to Purdue graduate students regarding Division C of AERA
- 2009 – 2011 Board Member, Arizona Association for the Gifted and Talented, Phoenix, Arizona
-Served as the Conference Co-Chair for the AAGT Annual State Conference, member of the scholarship committee, and member of the education committee.

AWARDS

- 2014 Feldhusen Doctoral Fellowship Award in Gifted Education
Purdue University, West Lafayette, IN
- 2013 Second Place, NAGC’s Graduate Student Research Gala
National Association for Gifted Children, Indianapolis, IN
- 2011 Passion and Commitment for Educational Excellence Award
Avondale Elementary School District, Avondale, Arizona
- 2003 Project Excellence Teacher Award
Warren County, Ohio
- 1999 – 2003 Full scholarship for M.A. in Teaching
Miami University, Oxford, Ohio
- 1995 -- 1996 Leadership and Service Award
Miami University, Oxford, Ohio

PUBLICATIONS

- McIntosh, J. (In Press). Seagulls, treasure maps, and scurvy: Sailing as a metaphor for differentiation. *Teaching for High Potential*. National Association for Gifted Children.

- Gentry, M., Peters, S., Pereira, N., McIntosh, J., & Fugate, C. (2015). *HOPE Teacher Rating Scale: Administration Manual*. Waco, TX: Prufrock Press.
- McIntosh, J. (2015). Gifted Program Evaluation: A Handbook for Administrators and Coordinators. Speirs Neumeister, K. & Burney, V. (2012), *Roeper Review*, 37:2, 111-112
- McIntosh, J. (2014). Curriculum compacting: Organized common sense. In M. Gentry (with K. A. Paul, J. McIntosh, C. M. Fugate, & E. Jen), *Total school cluster grouping and differentiation: A comprehensive, research-based plan for raising student achievement and improving teacher practices* (2nd ed., pp. 121-143). Waco, TX: Prufrock Press.
- McIntosh, J. (2014). Differentiation: Ceiling Demolition. In M. Gentry (with K. A. Paul, J. McIntosh, C. M. Fugate, & E. Jen), *Total school cluster grouping and differentiation: A comprehensive, research-based plan for raising student achievement and improving teacher practices* (2nd ed., pp. 99-120). Waco, TX: Prufrock Press.
- Patrick, H., Gentry, M., Moss, J., & McIntosh, J. (2014). Motivation and gifted adolescents. In Dixon & Moon's (Eds.) *The handbook of secondary gifted education*. Waco, TX: Prufrock Press.
- VanTassel-Baska, J., McIntosh, J., & Kearney, K. (2014). Secondary affective curriculum and instruction for gifted learners. In Dixon & Moon's (Eds.) *The handbook of secondary gifted education*. Waco, TX: Prufrock Press.
- McIntosh, J. (2013, Winter). Becoming a genuine bloomarian. *Teaching for High Potential*. National Association for Gifted Children.
- Menonno, A., Wackerly, A., & McIntosh, J. (2013). *Vets, pets, and me! Fat dogs and coughing horses elementary curriculum grades 1-3*. Purdue University and National Institutes of Health. Available from:

<http://www.vet.purdue.edu/engagement/files/documents/sepa/3rd%20Grade%20Curriculum%20Manual.pdf>

Ruhl, J., Veatch, J., & McIntosh, J. (2013). *Animal contributions to human health: Fat dogs and coughing horses high school curriculum, grade 9*. Purdue University and National Institutes of Health. Available from:

<http://www.vet.purdue.edu/engagement/files/documents/sepa/9th%20Grade%20Curriculum%20Manual.pdf>

McIntosh, J. (2011). The gap, the trap, and the high flyers flaps: A summary and analysis of three important studies focusing on excellence gaps in American education. *Gifted Children: An Electronic Journal of the AERA SIG Research on Giftedness and Talent*. 5 (1).

CONFERENCE PRESENTATIONS

McIntosh, J. (2014, November). *Viewing creativity as the map and not just the destination: A framework for using Wallace's four stages of creativity to develop curriculum*. National Association for Gifted Children Annual Conference. Baltimore, MD.

McIntosh, J. (2014, April). *A validation study of the HOPE Scale for use with 6th-12th grade Native American students*. Paper presented at the Annual Meeting of the American Educational Research Association. Philadelphia, PA.

McIntosh, J. (2013, November). *The General Pre-Assessment of Students (GPS) Instrument: A new tool for assessing students' prior knowledge and motivation*. National Association for Gifted Children Annual Conference. Indianapolis, IN.

McIntosh, J. (2013, October). *The Depth and Complexity Program Evaluation Tool (DC-PET): A new instrument for conducting internal evaluations of gifted education programs*. American Evaluation Association. Washington D.C.

McIntosh, J. (2013, May). *The social and emotional needs of gifted children*. 4th Annual National Leadership Summit; Identifying and Serving Gifted Native American Students. Lukachukai, AZ.

McIntosh, J. & Kim, J. (2013, April). *Designing, piloting, and evaluating concept based curriculum for high ability elementary students*. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA.

McIntosh, J. (2012, November) *Using a parent advisory council to create guardians of the gifted*. National Association for Gifted Children Annual Conference. Denver, Colorado.

McIntosh, J. (2012, May). *The recipe for an excelling student: Strategies for creating academic rigor for all students*. 3rd Annual National Leadership Summit; Identifying and Serving Gifted Native American Students. Ganado, Arizona.

McIntosh, J. (2011, November). *Dreaming with your eyes open: engaging gifted parents and children outside of the school day*. National Association for Gifted Children Annual Conference. New Orleans, Louisiana.

McIntosh, J. (2010, February). *Infusing creativity into academic content*. Arizona Association for Gifted and Talented Annual Conference. Phoenix, Arizona.

McIntosh, J. (2009, February). *Writing SMART goals for smart students*. Arizona Association for Gifted and Talented Annual Conference. Phoenix, Arizona.

PROFESSIONAL DEVELOPMENT PROVIDED TO SCHOOL DISTRICTS

McIntosh, J. (June, 2014). *Total school cluster grouping and differentiation*. Twin Lakes School District, IN.

McIntosh, J. (December, 2013). *The Kaplan depth and complexity model*. Lexington County School District #2, SC

McIntosh, J. (August, 2013). *Curriculum compacting*. Anoka-Hennepin District, MN

McIntosh, J. (April, 2013). *Curriculum compacting*. Beaverton School District, OR

PROFESSIONAL DEVELOPMENT PROVIDED TO AVONDALE SCHOOL DISTRICT

McIntosh, J. (2011). *Counseling the gifted child*. Avondale Elementary School District

- McIntosh, J. (2010). *Depth and complexity- Sandra Kaplan's model*. Avondale Elementary School District
- McIntosh, J. (2010). *Differentiation for middle school students*. Avondale Elementary School District
- McIntosh, J. (2010). *Using Bloom's taxonomy to write rigorous objectives*. Avondale Elementary School District
- McIntosh, J. (2009). *Curriculum compacting- helping our students buy back some of their time*. Avondale Elementary School District
- McIntosh, J. (2009). *Unlock your passions- implementing enrichment clusters*. Avondale Elementary School District
- McIntosh, J. (2009). *Bloom, Gardner, and beyond- how to meet the needs of all your students*. Avondale Elementary School District
- McIntosh, J. (2009). *What's different about differentiating for the gifted?*. Avondale Elementary School District
- McIntosh, J. (2008). *Thinking outside the box- critical and creative thinking skills for all*. Avondale Elementary School District
- McIntosh, J. (2008). *Academic rigor- having high expectations for students and staff*. Avondale Elementary School District
- McIntosh, J. (2008). *Tiered lessons- meeting the needs of all your students*. Avondale Elementary School District
- McIntosh, J. (2007). *Teaching gifted kids in the regular classroom*. Avondale Elementary School District
- McIntosh, J. (2007). *Active questioning- questioning makes the difference*. Avondale Elementary School District

TRAIN THE TRAINER AND CERTIFICATION PROGRAMS

- | | |
|---------------|---|
| February 2010 | Kids at Hope- Train the Trainer, Avondale, Arizona |
| January 2010 | SENG Parent Support Group Certification, Phoenix, Arizona |
| June 2009 | Thinking Maps: A Language for Learning – Train the Trainer, Scottsdale, Arizona |

February 2008	Structured English Immersion- 45 Clock Hour Certification, Avondale, Arizona
April 2006	Project Wild Certification, Savannah, Georgia
August 2003	Shared Inquiry Leader Workshop- Level I, Dayton, Ohio

PROFESSIONAL MEMBERSHIPS

American Educational Research Association (2012-2014)

American Evaluation Association (2012-2014)

Arizona Association for the Gifted (2009-2011, 2014-present)

National Association for Gifted Children (2008-present)

REFERENCES

Available Upon Request