Wayne State University Dissertations

1-1-2013

An Empirical Examination And Metaevaluation Of The Impact Evaluation Process

Anne Maechling Blake
Wayne State University,

Follow this and additional works at: http://digitalcommons.wayne.edu/oa_dissertations

AN EMPIRICAL EXAMINATION AND METAEVALUATION
OF THE IMPACT EVALUATION PROCESS

by

ANNE M. BLAKE

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2013

MAJOR: INSTRUCTIONAL TECHNOLOGY

Approved by:

_____
Advisor   Date

_____

_____

_____

DEDICATION

This dissertation is dedicated to my husband, Chris Blake,

who turned my life into a really fun adventure…

and to our children, Peter, Sarah, Maggie, and Lilly

who have shown us a glimpse of heaven.

ACKNOWLEDGEMENTS

I have had a lot of support from friends and family during the course of my graduate studies. It was gratifying to see that other people believed in my ability to accomplish this task even when I didn't completely believe in myself. However, it was due to the support of two men, in particular, that I was able to get this done. My husband, Chris, constantly encouraged me and picked up the slack at home. My advisor, Dr. James Moseley, has been a mentor to me in the truest sense of the word. For the past few years Dr. Moseley has provided me with countless academic opportunities, supported my research interests, pushed me to improve my work, identified gaps in my experience, and then sought ways to fill the gaps (a true Human Performance Technology professional!). His concern for his students, professionalism, and behavior as a gentleman have inspired me to pursue a career in academia.

I would like to thank the other members of my dissertation committee. Dr. Ingrid Guerra-Lopez was kind enough to let me study her Impact Evaluation Process model as part of this project. While working as her graduate research assistant, I learned a great deal. She has been an excellent role model. Dr. Ke Zhang has provided generous support and enthusiasm for my research topic. Throughout my doctoral studies, I have been able to count on the fact that Dr. Marcus Dickson will ask really hard questions, so each paragraph of this dissertation was written with the question "will this pass the Dr. Dickson test?" in mind, and for that I am grateful.

I would also like to thank other members of the Wayne State community for their support. Dr. Tim Spannaus has given me many professional opportunities over the past few years. Dr. Rita Richey was one of my first teachers in this program and inspired me to do my best work. Dr. Carol Diroff and Dr. Lynn Miller-Weitecha showed me how to be a good teacher.

Michele Norris was my cheerleader and organizer. Dr. Kelly Unger and Sara Kacin were the colleagues and friends whom I could call day or night for encouragement and laughter.

On a personal note, I would like to thank my parents, Joan and Phil Maechling, for their unwavering love and support. Their example of service to others and commitment to education set me on my life's path.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

Human Performance Technology (HPT) is a dynamic field that is constantly evolving and developing (Guerra-Lopez, 2008). It is dedicated to solving problems surrounding human performance in the workplace (Mager, 2006). The International Society for Performance Improvement (ISPI), which is the leading professional association in the field, defines Human Performance Technology (HPT) as:

> "a systematic approach to improving productivity and competence, (which) uses a set of methods and procedures -- and a strategy for solving problems -- for realizing opportunities related to the performance of people. More specifically, it is a process of selection, analysis, design, development, implementation, and evaluation of programs to most cost-effectively influence human behavior and accomplishment" (ISPI, 2012).

The field of HPT, which is increasingly referred to as the field of Performance Improvement (PI), has all of the characteristics identified by Finn (1960) as necessary to constitute a profession, including an intellectual technique, an application of this technique, a long period of training, a professional association of members, enforced standards and statements of ethics, and a body of intellectual theory. However, in spite of these robust attributes, scholars and practitioners have identified several ongoing challenges. One of the biggest challenges is that the field is based on an eclectic collection of elements, which draw upon the work of several related applied fields, such as organizational development, organizational psychology, human resources development, industrial engineering, and information technology (Marrelli, 2011,

Pershing, Lee & Cheng, 2008, Reiser & Dempsey, 2012). Pershing et al. (2008) identified several academic disciplines that have been influential in the development of the field, including systems theory, cognitive science, psychology, economics, and philosophy. Pershing (2006) believes that this eclectic nature has a downside. He states that "drawing upon the principles and theories of numerous academic disciplines and other fields contributes to a lack of clarity for HPT" (p. 29).

An additional concern is that practitioners and academics within the field continue to encounter skeptics who claim that PI should be called a "craft" rather than a profession because professional practice is not solidly based on empirical research (Kaufman & Clark, 1999). There is a concern that the field lacks a shared understanding of "key concepts, principles, and theories and how they are applied in practice through models and methods" (Marrelli, 2011, p. 6). In order to advance as a profession and a field of study, performance improvement practitioners and scholars need to develop common terminology, methods, and models (Marrelli, 2011).

Finally, some leaders within the field of performance improvement believe that the empirical foundations of the field have not kept pace with practice and continue to call for an increase in targeted research activity (Stolovitch, 2000; Sugrue & Stolovitch, 2000). Klein's (2002) study found that "appeals for empirical research are going unheeded" (p. 104).

The Performance Improvement Process

In order to understand the potential gaps in the existing research base for the field of PI, it is necessary to understand the PI process itself. A graphic depiction of the Performance Improvement/HPT process appears in Figure 1.

Figure 1. Performance Improvement/HPT Model

Performance Improvement/HPT Model
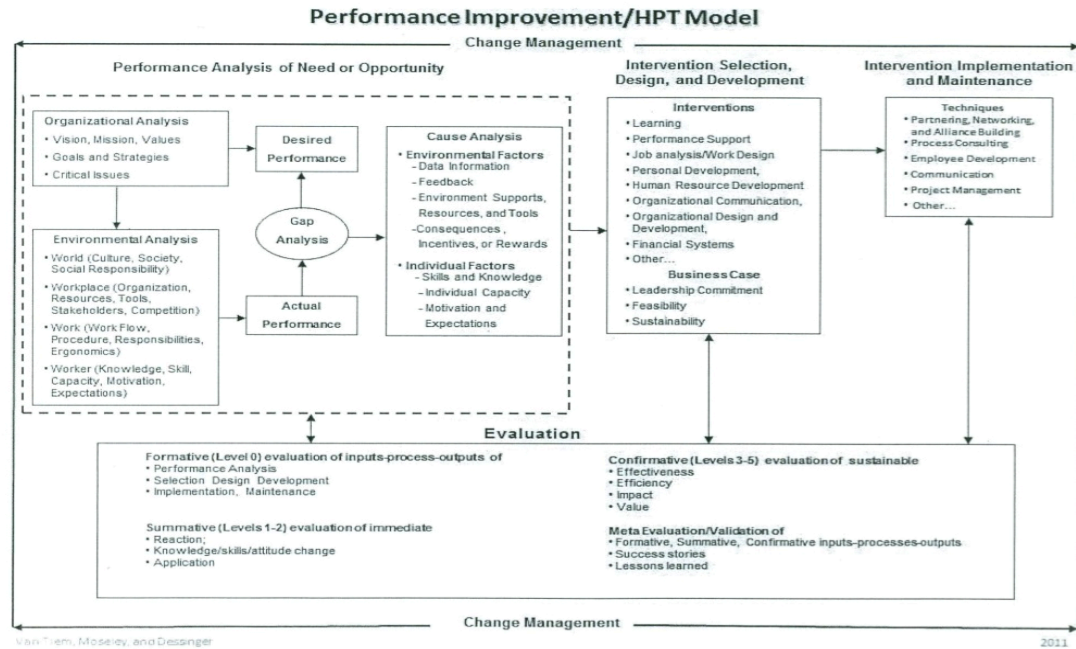
Van Tiem, Moseley, & Dessinger (2012).Used with permission.

As depicted in the model, the first step in the PI process is to conduct a performance analysis to identify gaps between what is actually happening in the workplace and what should be happening now or in the near future. A thorough performance analysis includes an organizational analysis, an environmental analysis, a gap analysis and a cause analysis to identify the root causes of all identified performance deficiencies. The next step is to select, design, and develop appropriate interventions that will enable people to perform at the desired level. When this is complete, the interventions are implemented. Throughout the process, and at its completion, the user is encouraged to conduct thorough and systematic evaluation (Brethower,

2012). The concept of change management surrounds the entire performance improvement effort. After all, change is at the heart of every improvement effort.

Key contributors to the PI theoretical base include researchers who work primarily within the field as well as scholars from other disciplines whose theories are relevant to performance improvement. Empirical research has been conducted and models have been developed based on the PI process in its entirety as well as on individual elements of the process. For example, Roger Kaufman's Organizational Elements Model (2006) focuses on needs assessment and analysis. Thomas Gilbert's Behavioral Engineering Model (1978) is particularly geared towards intervention selection. Kirkpatrick's Four Levels of Evaluation (1959) is a well-known model that has been primarily used as part of the evaluation component of the PI process.

Evaluation

While evaluation is an element of the PI process, it is also a field of study that has been widely researched in its own right. In fact, the field of evaluation is a good example of a mature professional discipline, since it has all of the characteristics identified by Finn (1960) and fulfills Marrelli's (2011) requirement that practitioners and scholars have common terminology, methods, and models. While it has also grown out of different intellectual traditions and approaches to practice, scholars and practitioners have done an excellent job of developing common terminology and professional standards. Under the guidance of the leading professional association, The American Evaluation Association (2012), the field has agreed upon *Guiding Principles for Evaluators* (Fitzpatrick, Sanders & Worthen, 2011). These principles "provide ethical guidance for evaluators in their everyday practice" as well as "inform us as to ethical and appropriate ways for evaluations to be conducted" (Fitzpatrick et al., 2011, p. 82). The Joint Committee on Standards for Educational Evaluation maintains an extensive set of *Program*

Evaluation Standards (Yarbrough, Shulha, Hopson & Caruthers, 2011), which are "designed to assist evaluators and consumers in judging the quality of a particular evaluation" (Fitzpatrick et al., 2011, p. 82). Scholars within the field have developed several similar frameworks of approaches to evaluation that they consider to be the core of the field (Christie & Alkin, 2012; Fitzpatrick et al., 2011; Stufflebeam 2001a).

Evaluation is a critical component of the PI process, rather than something that is done as a one-time event or as an afterthought (Van Tiem, Moseley & Dessinger, 2012). In fact, several important models of evaluation have developed from within the field of PI. These include: Kirkpatrick's Four Levels of Evaluation (1994), Brinkerhoff's Success Case Method (2003), and Dessinger & Moseley's (2004) Confirmative Evaluation, among others. Some of the leaders within the field of evaluation are also identified as key contributors to the field of PI, including Michael Scriven and Daniel Stufflebeam. Unfortunately, a search of the evaluation literature indicates that PI scholars are rarely mentioned in evaluation's scholarly journals. The field of PI would benefit from positioning its own scholars' models within the theoretical framework of the field of evaluation.

The most recent evaluation model within the field of PI is the Impact Evaluation Process (IEP) (Guerra-Lopez, 2007a). The model is discussed in greater detail in the following section. The model was heavily influenced by Kaufman's (2006) work on needs assessment and is based on his identification of three basic levels of results: mega, macro, and micro.

Guerra-Lopez (2008) has identified several strengths and limitations to the IEP. Strengths include a strong focus on aligning performance with the ultimate desired impact and providing detailed guidance in conducting an evaluation in its entirety. A limitation of the model is the fact that little research has been conducted using the framework. Several conceptual pieces have been

published within HPT literature (Guerra-Lopez, 2007b, 2007c, 2007d, 2008, 2010, 2012), and a recent research article has been published within the evaluation literature (Guerra-Lopez & Toker, 2012). However, additional research is needed to empirically evaluate the model, assess its feasibility, and identify its position within the theoretical framework of the field of evaluation.

Statement of the Problem

The field of Performance Improvement continues to face several challenges, including a concern that professional practice is not always based on research and that the empirical foundations of the field have not kept pace with practice. As a result, there is a need for targeted research activity to clarify and cement elements of the field within their broader intellectual traditions and to clarify the connections between theory and practice.

The Guerra-Lopez IEP (2007a) is a relatively new and promising model within the field, and is increasingly being used in graduate evaluation courses and professional practice. However, it would benefit from additional empirical research in order to bridge the gap between theory and practice and to anchor its position within the field of evaluation.

Purpose of the Study and Research Questions

This research focuses on a qualitative case study evaluation using the IEP as well as an empirical evaluation and metaevaluation of the IEP. The study stems from a constructivist theoretical perspective, with the goal of understanding the meaning that the IEP has for the people who use it (Crotty, 1998; Merriam, 2009), and to use this understanding to draw conclusions about the model's effectiveness.

During the first phase of the study, a program evaluation was conducted using the Guerra-Lopez IEP. The setting for the program evaluation is a one-to-one (1:1) technology program in a secondary school. The evaluation of the technology program provides a basis to

make preliminary judgments about the effectiveness of the IEP for evaluating educational programs.

When the initial program evaluation was complete, an empirical examination and metaevaluation of the Guerra-Lopez IEP was conducted. In order to triangulate the results of the metaevaluation, and to reduce the potential for evaluator bias, the metaevaluation was conducted by all three of the groups identified by Fitzpatrick et al. (2011) as appropriate to conduct metaevaluations: the evaluation consumers, the person who conducted the evaluation, and an outside expert evaluator. The metaevaluation was conducted using Stufflebeam's *Program Evaluations Metaevaluation Checklist* (2011). The checklist is based on the Joint Committee's *Program Evaluation Standards* (Yarbrough et al., 2011) and will be discussed in greater detail in the next section. The researcher has received permission to use this checklist from Dr. Daniel Stufflebeam and his co-author, Dr. Chris Coryn, in its newly revised form that is in press. A copy of this approval appears in Appendix A.

The purpose of the study is threefold: to identify where the IEP fits within the body of evaluation theory, to assess the effectiveness of the Guerra-Lopez IEP as a tool to evaluate educational programs, and to empirically examine the IEP from multiple perspectives, primarily using Stufflebeam's (2011) *Metaevaluation Checklist*. The *Checklist* is based on standards in five areas of concern: utility, feasibility, propriety, accuracy, and accountability.

The research questions that follow from this statement of purpose are:

Research Question 1: Where does the Guerra-Lopez IEP fit into the body of evaluation research?

Research Question 2: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Utility?

Research Question 3: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Feasibility?

Research Question 4: To what extent does the IEP meet the Joint Committee Program Evaluation requirements for Propriety?

Research Question 5: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accuracy?

Research Question 6: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accountability?

Research Question 7: How effective is the IEP for evaluating educational programs?

Rationale and Significance of the Study

The rationale for this research study stems from an interest in helping the field of performance improvement synthesize its eclectic roots into a cohesive and empirically solid field of study and practice. The only way to address the nagging criticism that practice is not solidly based on empirical research is to make a concerted effort to match research to practice.

The significance of the study is that it will contribute to the domain's knowledge base by empirically evaluating one of its most recently developed and published models. Further, this study contributes to the field of evaluation because it tests the IEP against the accepted standards of the field of evaluation. It adds to the metaevaluation research base by providing the first empirical research based on Stufflebeam's revised (2011) *Program Evaluations Metaevaluation Checklist.*

Theoretical Constructs and Models

The theoretical construct of metaevaluation is at the heart of this study. In addition, several models are central to the proposed research, including Kaufman's Organizational

Elements Model (2006) and the Guerra-Lopez Impact Evaluation Process. Finally, the Stufflebeam (2011) *Metaevaluation Checklist* is a vital part of the research and should be clearly understood at the outset.

Metaevaluation

Metaevaluation is frequently described as "an evaluation of an evaluation" (Fitzpatrick et al., 2011; Henry & Mark, 2003; Stufflebeam, 2001b). Metaevaluations are systematic reviews of evaluations that help to determine the quality of their processes and findings (Cooksy & Caracelli, 2009). "Evaluators need metaevaluations to assure the quality of their evaluations, provide direction for improving individual studies as well as their developing evaluation approaches, and earn and maintain credibility for their services among both clients and other evaluators" (Stufflebeam, 2001b, p. 184). As such, a metaevaluation of the Guerra-Lopez IEP is a valuable step in establishing the model within the field.

Frequently, metaevaluations are conducted on individual evaluations to control for bias. (Fitzpatrick et al., 2011). However, Nilsson and Hogben (1983) point out that metaevaluation does not only refer to evaluations of particular studies, but also to the evaluation of the practice and function of evaluation itself. This study included both types of metaevaluations since it consisted of a metaevaluation of an individual evaluation as well as an analysis of the design. Fitzpatrick et al. (2011) suggest the following steps for conducting a metaevaluation of an evaluation design: (1) Prepare a copy of the design in a form that is ready for review, (2) Clearly identify who will do the evaluation, (3) Verify that approval has been given to evaluate the design, (4) Apply the appropriate standards to the evaluation design, and (5) Judge the adequacy of the evaluation design.

Fitzpatrick et al. (2011) identify three groups that are appropriate to conduct metaevaluations. First, metaevaluations can be conducted by the original evaluator. This method has the most potential to succumb to bias, and is the least recommended method of metaevaluation. The second potential metaevaluation method is to have it conducted by the evaluation's consumers. Fitzpatrick et al. (2004) state that the success of this approach is heavily dependent on the consumer's technical ability to judge how well the evaluation meets the standards identified by the Joint Committee. The authors point out that these standards do not require technical training, and that it is feasible for a client to apply the criteria effectively.

The final metaevaluation method is to have it conducted by expert evaluators. All else being equal, this seems like the best approach (Fitzpatrick et al., 2011). For example, Stufflebeam and Coryn (2012) defend their method of metaevaluating nine evaluation designs by describing in detail their experience and expertise as evaluators. The current study employed all three metaevaluative groups.

Kaufman's Organizational Elements Model

The development of the Guerra-Lopez Impact Evaluation Process (IEP) was heavily influenced by Roger Kaufman's (2006) work with needs assessment. Kaufman differentiates among three basic levels of results: strategic, tactical, and operational. Strategic results are long term and ultimately benefit society as well as the client. Tactical results are shorter-term and represent organization-wide achievements. Operational results are "building-block objectives" (Guerra-Lopez, 2008, p. 82) that help the organization reach its mission when they are combined. In addition, Kaufman makes a distinction between means and ends, where the ends manifest themselves in results. The evaluand, the thing that is being evaluated, is always a means to an end, where the end manifests itself at one of the levels of results listed above. For example, in the

case of the current study, the technology program (the evaluand) is a means to an end at the secondary school where it was implemented. The end can exist at the strategic level, such as "Creating Women Who Make a Difference", which is the school's motto, or at a lower level, such as "providing students with 21st Century skills". An illustration of Kaufman's model is shown below in Table 1.

Table 1. Kaufman's Organizational Elements Model

Name of Organizational Element	Brief Description and Level of Focus	Level of Planning and Results
Mega	Results and their consequences for external clients and society (shared vision)	Strategic
Macro	The results and their consequences for what an organization can or does deliver outside of itself	Tactical
Micro	The results and their consequences for individuals and small groups within the organization	Operational
Process	Means, programs, projects, activities, methods, techniques	
Input	Human, capital, and physical resources; existing rules, regulations, policies, laws	

Kaufman, R.A. (2006).

Guerra-Lopez Impact Evaluation Process

Needs assessment and evaluation are compatible concepts. A needs assessment helps to "create the future" (Guerra-Lopez, 2007a, p. 7) by articulating a performance-based vision and relevant objectives, as well as identifying the path to reach the vision. An evaluation, on the other hand, determines whether or not the organization is on the right track towards "reaching the future it set out to create during the needs assessment process" (Guerra-Lopez, 2007a, p. 7). The

Guerra-Lopez IEP is a natural progression from Kaufman's model, since it is based on the same organizational elements and levels of results.

A key principle of the IEP is that "everything we do is aligned to some greater purpose, whether we are conscious of it or not and whether we are aligning it well or not" (Guerra-Lopez, 2008, p. 81-82.). The model is intuitively appealing within the field of PI because it builds on Kaufman's model for conducting needs assessment. With a common conceptual framework and terminology, it is easy to see a natural progression from a well-designed needs assessment, through intervention selection, development and implementation of interventions, and ultimately to evaluation.

The model consists of seven elements. Although they are described in sequence, they should be considered reiteratively. The process consists of: (1) Identifying stakeholders and expectations, (2) Determining key decisions and objectives, (3) Deriving measurable indicators, (4) identifying data sources, (5) Selecting data collection methods, (6) Selecting data analysis tools, and (7) Communicating results and recommendations. The Impact Evaluation Process is depicted in graphic form in Figure 2 below. Each step is described in more detail in Table 2 below:

Figure 2. The Impact Evaluation Process
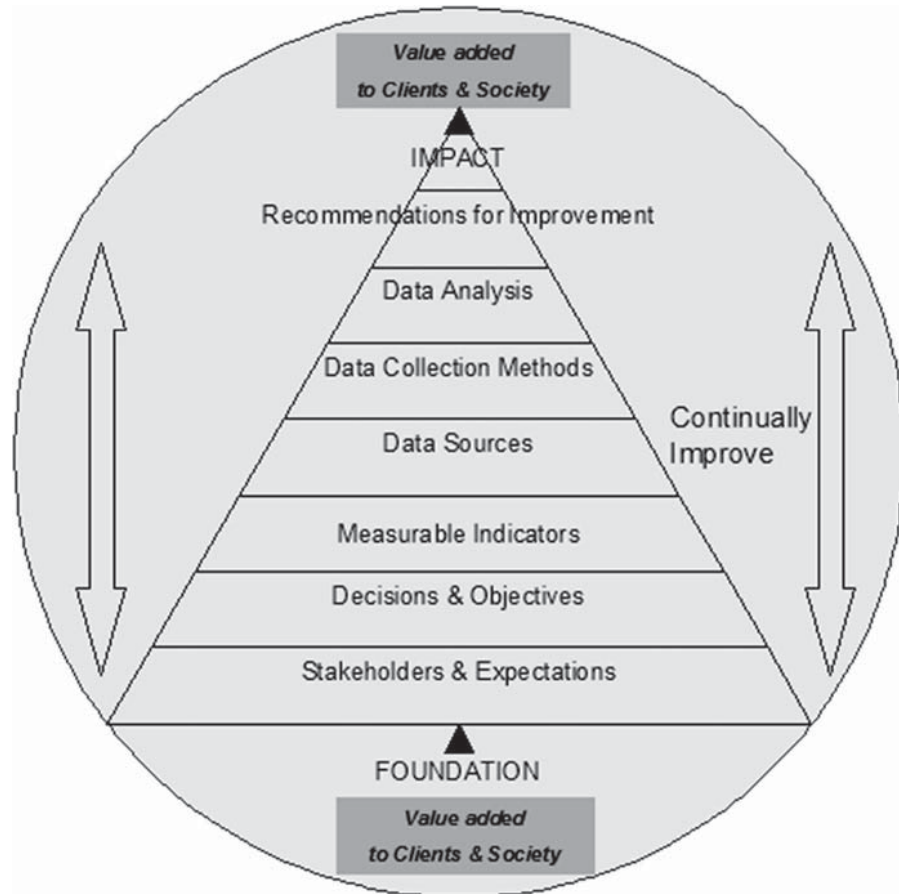
Guerra-Lopez (2008). Used with permission.

Table 2: The Impact Evaluation Process

Evaluation Step	Description
1. Identify stakeholders and expectations	• The evaluator identifies key stakeholders involved. Stakeholder groups include those who will be making decisions either throughout the evaluation process, or directly as a result of the evaluation findings. This will include those with authority to make critical decisions (often those who finance the evaluation project), but should also include those who will be affected by the evaluation • The driving question for identifying stakeholders is "who is/could be either impacted by the evaluation, or could potentially impact the evaluation in a meaningful way?"

	• While not every single stakeholder must be directly involved as part of the evaluation project team, each group should be represented so that project remains aligned with expectations of each group
	• Clearly identify each group's expectations. Typical expectations include: specific questions to be answered, time frames, final report content, data access, what is expected of stakeholders and the evaluator
	• The evaluator identifies why stakeholders want to conduct the evaluation and how they will define a "successful" evaluation
2. Determine key decisions and objectives	• The evaluator asks stakeholders to articulate what types of decisions will be made as a result of the findings. This discussion must include key goals and objectives internal and external to the organization
	• The evaluator helps stakeholders articulate and agree on these objectives and decisions points
	• Evaluation questions are developed based on the agreed upon objectives and decision points
	• Evaluation questions are aligned with desired organizational results at all levels
3. Derive measurable indicators	• Sound decisions are made on the basis or relevant, reliable, valid, and complete data related to desired results, and the related questions we want to answer. Therefore, the heart of the evaluation plan will be to gather data required to answer the questions that guide the inquiry
	• Performance indicators are observable phenomena that are linked to something that is not directly observed and can provide information that will answer an evaluation question
	• The evaluator will develop a list of performance indicators that point towards each evaluation question that was identified in the previous step
	• The IEP highlights that it is critical to measure both lagging indicators (measures of ultimate results) and leading indicators (key performance metrics that are tracked on a regular basis for purposes of monitoring, feedback, and continual improvement)

4. Identify data sources	• With a list of specific indicators for which to collect data, the evaluator determines where to find the data. The required data point to the appropriate source. It is likely that most of the required data will be available within the organization • Excellent sources include: strategic plans, annual reports, project plans, consulting studies, performance reports, the internet, and other technologies that will allow the user to access reports, documents, databases, experts, and other sources. Many companies, government agencies and research institutions publish official studies and reports that could also prove to be valuable
5. Select data collection instruments	• The right data collection methods and tools are a function of the data you are seeking. Likewise, the data you collect is a function of the methods you select • There is extensive literature about data collection methods. Selection should be made based on pros and cons, specifically with regards to important criteria such as appropriateness of the instrument for the required data, time characteristics of sample, comprehensiveness of tool, previous experience with tools that are being considered, and feasibility among others
6. Select data analysis tools	• Data analysis is more than number crunching. It is the organization of information to discover patterns and fortify arguments used to support conclusions or evaluative claims that result from your evaluation study • The evaluator summarizes large volumes of data into a manageable and meaningful format that can quickly communicate its meaning
7. Communicate results and recommendations	• The importance of effective communication cannot be overstated. A rigorous evaluation does not speak for itself. Communicating with key stakeholders throughout the evaluation process keeps them aware of what you are doing and why, which in turn increases the amount of trust they place in you and your efforts • The evaluator develops the report based an understanding

> of the audience that will receive it and adjusts language and format accordingly
> - The evaluator will clearly articulate what needs to be done as a result of the evaluation as well as who is responsible for implementation and an explanation of *how* to implement the recommendations
> - The evaluator will make clear distinctions and linkages among recommendations, interpretations, findings, and the analyzed data

Based on Guerra-Lopez & Toker, 2012.

Stufflebeam's Program Evaluation Metaevaluation Checklist

Dr. Daniel Stufflebeam created the *Program Evaluations Metaevaluation Checklist* (2011) based on his participation in the creation of the Joint Committee's standards for evaluations of programs and personnel as well as his efforts to research and apply the standards (Stufflebeam & Coryn, 2012). The purpose of the *Checklist* is to assess program evaluations against professionally defined requirements for solid evaluations. The checklist consists of five parts, and includes instructions for using the checklist, steps for preparing to conduct the metaevaluation, a recommended format for describing the subject evaluation, the core element that rates the subject evaluation against the 30 Joint Committee *Standards*, tables for analyzing and summarizing information, and a format for preparing the final metaevaluation report.

The stakeholders who participated in the metaevaluation portion of this study were asked to identify whether or not the evaluation process addressed each of the Joint Committee's 30 *Standards*. For each of the standards, there are six "checkpoints" (Stufflebeam 2011, p. 3) that were addressed. The standards and checkpoints are listed in Table 3 below.

Table 3. Program Evaluation Metaevaluation Checklist

Section	Standard	Checkpoint
Utility	U1.	U1.1 Engage evaluators who possess the needed knowledge,

Section	Standard	Checkpoint
	Evaluator Credibility	skills, experience, and professional credentials
		U1.2 Engage evaluators whose evaluation qualifications, communication skills, and methodological approach are a good fit to the stakeholders' situation and needs
		U1.3 Engage evaluators who are appropriately sensitive and responsive to issues of gender, socioeconomic status, race, language, and culture
		U1.4 Engage evaluators who build good working relationships, and listen, observe, clarify, and attend appropriately to stakeholders' criticisms and suggestions
		U1.5 Engage evaluators who have a record of keeping evaluations moving forward while effectively addressing evaluation user's information needs
		U1.6 Give stakeholders information on the evaluation plan's technical quality and practicality, e.g., as assessed by an independent evaluation expert
	U2. Attention to Stakeholders	U2.1 Clearly identify and arrange for ongoing interaction with the evaluation client
		U2.2 Identify and arrange for appropriate exchange with the other right-to-know audiences, including, among others, the program's authority figures, implementers, beneficiaries, and funders
		U2.3 Search out and invite input from groups or communities whose perspectives are typically excluded, especially stakeholders who might be hindered by the evaluation
		U2.4 Help stakeholders understand the evaluation's boundaries and purposes and engage them to uncover assumptions, interests, values, behaviors, and concerns regarding the program
		U2.5 Determine how stakeholders intend to use the evaluation's findings
		U2.6 Involve and inform stakeholders about the evaluation's

Section	Standard	Checkpoint
		progress and findings throughout the process, as appropriate
	U3. Negotiated Purpose	U3.1 Identify the client's stated purposes for the evaluation
		U3.2 Engage the client and stakeholders to weigh stated evaluation purposes – e.g., against their perceptions of dilemmas, quandaries, and desired evaluation outcomes – and to embrace evaluation's bottom line goal of assessing value, e.g., a program's worth, merit, or significance
		U3.3 Help the client group consider possible alternative evaluation purposes, e.g., program planning, development, management, and improvement: program documentation and accountability; and judging the program's quality, impacts, and worth
		U3.4 Engage the client to clarify and prioritize the evaluation's purposes using appropriate tools such as needs assessments and logic models
		U3.5 Provide for engaging the client group periodically to revisit and, as appropriate, update the evaluation's purpose
		U3.6 Assure that initial and updated evaluation purposes are communicated to the full range of stakeholders
	U4. Explicit Values	U4.1 Make clear the evaluator's commitment to certain, relevant values, e.g., an evaluation's utility, feasibility, propriety, accuracy, and accountability and a program's equity, fairness, excellence, effectiveness, safety, efficiency, fiscal accountability, legality, and freedom from fraud, waste, and abuse
		U4.2 Engage the client and program stakeholders in an effective process of values clarification, which may include examining the needs of targeted program beneficiaries, the basis for program goals, and the rationale for defined evaluation purposes
		U4.3 Assist the client group to air and discuss their common and discrepant views of what values and purposes should guide the program evaluation

Section	Standard	Checkpoint
		U4.4 Acknowledge and show respect for stakeholders' possibly diverse perspectives on value matters, e.g., by assisting them to seek consensus or at least reach an accommodation regarding possible alternative interpretations of findings against different values
		U4.5 Clarify the values that will undergird the evaluation, taking account of client, stakeholder, and evaluator positions on this matter
		U4.6 Act to ensure that the client and full range of stakeholders understand and respect the values that will guide the collection, analysis, and interpretation of the evaluation's information
	U5. Relevant Information	U5.1 Interview stakeholders to determine their different perspectives, information needs, and views of what constitutes credible, acceptable information
		U5.2 Plan to obtain sufficient information to address the client group's most important information needs
		U5.3 Assess and adapt the information collection plan to assure adequate scope for assessing the program's value, e.g., its worth, merit, or significance
		U5.4 Assure that the obtained information will address and keep within the boundaries of the evaluation's stated purposes and key questions
		U5.5 Allocate time and resources to collecting different parts of the needed information in consideration of their differential importance
		U5.6 Allow flexibility during the evaluation process for revising the information collection plan pursuant to emergence of new, legitimate information needs
	U6. Meaningful Processes and	U6.1 Budget evaluation time and resources to allow for meaningful exchange with stakeholders throughout the evaluation process

Section	Standard	Checkpoint
	Products	U6.2 Engage the full ranges of stakeholders to assess the original evaluation plan's meaningfulness for their intended uses
		U6.3 During the evaluation process, regularly visit with stakeholders to assess their evaluation needs and expectations, also, as appropriate, to obtain their assistance in executing the evaluation plan
		U6.4 Regularly obtain stakeholders' reactions to the meaningfulness of evaluation procedures and processes
		U6.5 Invite stakeholders to react to and discuss the accuracy, clarity, and meaningfulness of evaluation reports
		U6.6 As appropriate, adapt evaluation procedures, processes, and reports to assure that they meaningfully address stakeholder needs
	U7. Timeliness and Appropriate Communi-cation and Reporting	U7.1 Plan to deliver evaluation feedback pursuant to the client group's projection of when they will need reports, but allow flexibility for responding to changes in the program's timeline and needs
		U7.2 Plan, as appropriate, to give stakeholders access to important information as it emerges
		U7.3 Employ reporting formats and media that accommodate the characteristics and serve the needs of the different audiences
		U7.4 Determine how much technical detail to report by identifying and taking account of the audience's technical background and expectations
		U7.5 Plan and budget evaluation follow-up activities so that the evaluator can assist the client group to interpret and make effective use of the final evaluation report
		U7.6 Pursuant to the above checkpoints, formalize expectations for communicating and reporting to the sponsor and stakeholders in the evaluation contract
	U8.	U8.1 Identify the stakeholders' formal and informal communication mechanisms that connect stakeholders and, as

Section	Standard	Checkpoint
	Concern for Consequences and Influence	appropriate, channel evaluation findings through these mechanisms
		U8.2 Be vigilant and proactive in identifying and appropriately communicating with stakeholders who appear to be sabotaging the evaluation and, as necessary, counteract the sabotage
		U8.3 Plan to meet, as appropriate, with stakeholders to help them apply findings in ways that are logical, meaningful, ethical, effective, and transparent
		U8.4 In discussing evaluation findings with the client group stress the importance of applying the findings in accordance with the evaluation's negotiated purposes
		U8.5 Be vigilant to identify, prevent, or appropriately address any misuses of evaluation findings
		U8.6 Follow up evaluation reports to determine if and how stakeholders applied the findings
Feasibility	F1. Project Management	F1.1 Ground management of the evaluation in knowledge of the stakeholders' environment and needs and the evaluation's purpose
		F1.2 Prepare a formal management plan including, e.g., the evaluation's goals, procedures, assignments, communication, reporting, schedule, budget, monitoring arrangements, risk management arrangements, and accounting procedures
		F1.3 Recruit evaluation staff members who collectively have knowledge, skills, and experience required to execute, explain, monitor, and maintain rigor, viability, and credibility in the evaluation process
		F1.4 Involve and regularly inform an appropriate range of stakeholders
		F1.5 Systematically oversee and document the evaluation's activities and expenditures
		F1.6 Periodically review the evaluation's progress and, as

Section	Standard	Checkpoint
		appropriate, update the evaluation plan and procedures
	F2. Practical Procedures	F2.1 Assess and confirm the program's evaluability before deciding to proceed with the evaluation
		F2.2 Employ procedures that fit well within the program and its environment
		F2.3 Assure that the selected procedures take account of and equitably accommodate the characteristics and needs of diverse stakeholders
		F2.4 Obtain relevant insider knowledge and incorporate it into the data collection process
		F2.5 Make efficient use of existing information and avoid needless duplication in collecting data
		F2.6 Conduct the evaluation so as to minimize disruption to the program
	F3. Contextual Viability	F3.1 Investigate the program's cultural, political, and economic contexts by reviewing such items as the program's funding proposal, budget documents, organizational charts, reports, and news media accounts and by interviewing such stakeholders as the program's funder, policy board members, director, staff, recipients, and area residents
		F3.2 Take into account the interests and needs of stakeholders in the process of designing, contracting for, and staffing the evaluation
		F3.3 Enlist stakeholder and interest group support through such means as regular exchange with a review panel composed of a representative group of stakeholders
		F3.4 Practice even-handedness and responsiveness in relating to all stakeholders, e.g., in the composition of focus groups
		F3.5 Avert or identify and counteract attempts to bias or misapply the findings
		F3.6 Provide appropriate mechanisms for stakeholders to remain informed about the evaluation's progress and findings, such as an

Section	Standard	Checkpoint
		evaluation project website, an evaluation newsletter, targeted reports, and a telephone response line
	F4. Resource Use	F4.1 Negotiate a budget--ensuring that the contracted evaluation work can be completed efficiently and effectively—to include the needed funds and the necessary in-kind support and cooperation of program personnel
		F4.2 Balance effectiveness and efficiency in resource use to help ensure that the evaluation will be worth its costs and that sponsors will get their money's worth
		F4.3 Use resources carefully with as little waste as possible
		F4.4 Utilize existing data, systems, and services when they are well aligned with the evaluation's purposes
		F4.5 Document the evaluation's costs, including time, human resources, expenditures, infrastructure support, and foregone opportunities
		F4.6 Document the evaluation's benefits, including contributions to program improvement, future funding, better informed stakeholders, and dissemination of effective services
Propriety	P1. Responsive and Inclusive Orientation	P1.1 Acquire and take account of knowledge of the program environment's history, significant events, culture, and other factors affecting the program and its evaluation
		P1.2 Identify stakeholders broadly, gather useful information from them, and include them, as appropriate, in decisions about the evaluation's purposes, questions, and design
		P1.3 Engage and serve the full range of stakeholders in an even-handed manner, regardless of their politics, personal characteristics, status, or power
		P1.4 Design and schedule the evaluation to provide multiple opportunities for stakeholders to be involved, contribute, and be heard throughout the evaluation process
		P1.5 Be open to and thoughtfully consider stakeholders' contradictory views, interests, and beliefs regarding the program's prior history, goals, status, achievements, and significance

Section	Standard	Checkpoint
		P1.6 Avert or counteract moves by powerful stakeholders to dominate in determining evaluation purposes, questions, and procedures and interpreting outcomes
	P2. Formal Agreements	P2.1 Negotiate evaluation-related obligations, with the client, including what is to be done, how, by whom, when, and at what cost
		P2.2 Make ethical, legal, and professional stipulations and obligations explicit and binding regarding such evaluation matters as evaluation purposes and questions, confidentiality/anonymity of data, editorial authority, release of reports, evaluation follow-up activities, cooperation of program staff, funds and in-kind resources, and provision for a metaevaluation
		P2.3 Employ the contract negotiation process to strengthen trust in communications through stakeholder consultation and, unless restricted by laws or regulations, allowing stakeholders to review the printed agreement
		P2.4 Ensure that formal evaluation agreements conform to federal, tribal, state, or local requirements, statutes, and regulations
		P2.5 Employ negotiated agreements to monitor, track, and assure effective implementation of specific duties and responsibilities
		P2.6 Revisit evaluation agreements over time and negotiate revisions as appropriate
	P3. Human Rights and Respect	P3.1 Adhere to applicable federal, state, local, and tribal regulations and requirements, including those of Institutional Review Boards, local/tribal constituencies, and ethics committees that authorize consent for conduct of research and evaluation studies
		P3.2 Take the initiative to learn, understand, and respect stakeholders' cultural and social backgrounds, local mores, and institutional protocols
		P3.3 Make clear to the client and stakeholders the evaluator's ethical principles and codes of professional conduct, including the standards of the Joint Committee on Standards for Educational

Section	Standard	Checkpoint
		Evaluation
		P3.4 Institute and observe rules, protocols, and procedures to ensure that all evaluation team members will develop rapport with and consistently manifest respect for stakeholders and protect their rights
		P3.5 Make stakeholders aware of their rights to participate, withdraw, or challenge decisions that are being made at any time during the evaluation process
		P3.6 Monitor the interactions of evaluation team members and stakeholders and act as appropriate to ensure continuing, functional, and respectful communication and interpersonal contacts throughout the evaluation
	P4. Clarity and Fairness	P4.1 Develop and communicate rules that assure fairness and transparency in deciding how best to allocate available evaluation resources to address the possible competing needs of different evaluation stakeholders
		P4.2 Assure that the evaluation's purposes, questions, procedures, and findings are transparent and accessible by all right-to-know audiences
		P4.3 Communicate to all stakeholders the evaluation's purposes, questions, and procedures and their underlying rationale
		P4.4 Make clear and justify any differential valuing of any stakeholders' evaluation needs over those of others
		P4.5 Carefully monitor and communicate to all right-to-know audiences the evaluation's progress and findings and do so throughout all phases of the evaluation
		P4.6 Scrupulously avoid and prevent any evaluation-related action that is unfair to anyone
	P5. Transparency and Disclosure	P5.1 Identify and disclose to all stakeholders the legal and contractual constraints under which the evaluation's information can be released and disseminated
		P5.2 Maintain open lines of communication with and be accessible to, at least representatives of, the full range of

Section	Standard	Checkpoint
		stakeholders throughout the evaluation, so they can obtain the information which they are authorized to review
		P5.3 Before releasing the evaluation's findings, inform each intended recipient of the evaluation's policies— regarding such matters as right-to-know audiences, human rights, confidentiality, and privacy—and, as appropriate, acquire her or his written agreement to comply with these policies
		P5.4 Provide all stakeholders access to a full description and assessment of the program, e.g., its targeted and actual beneficiaries; its aims, structure, staff, process, and costs; and its strengths, weaknesses, and side effects
		P5.5 Provide all stakeholders with information on the evaluation's conclusions and limitations
		P5.6 Provide all right-to-know audiences with access to information on the evaluation's sources of monitory and in-kind support
	P6. Conflicts of Interest	P6.1 Throughout the evaluation process search for potential, suspected, or actual conflicts of interest
		P6.2 Search for conflicts involving a wide range of persons and groups, e.g., those associated with the client, the program's financial sponsor, program recipients, area residents, the evaluator, and other stakeholders
		P6.3 Search for various kinds of conflicting interests, including prospects for financial gains or losses, competing program goals, alternative program procedures, alternative evaluation approaches, and alternative bases for interpreting findings
		P6.4 Take appropriate steps to manage identified conflicts so that the evaluation maintains integrity and high quality
		P6.5 Attend to conflicts of interest through effective communication with the client and other pertinent parties and in a spirit of mutual and deliberate understanding and learning
		P6.6 Document and report identified conflicts of interest, how they were addressed, and how they affected the evaluation's

Section	Standard	Checkpoint
		soundness
	P7. Fiscal Responsibility	P7.1 Plan and obtain approval of the evaluation budget before beginning evaluation implementation
		P7.2 Be frugal in expending evaluation resources
		P7.3 Employ professionally accepted accounting and auditing practices
		P7.4 Maintain accurate and clear fiscal records detailing exact expenditures, including adequate personnel records concerning job allocations and time spent on the job
		P7.5 Make accounting records and audit reports available for oversight purposes and inspection by stakeholders
		P7.6 Plan for and obtain appropriate approval for needed budgetary modifications over time or because of unexpected problems
Accuracy	A1. Justified Conclusions and Decisions	A1.1 Address each contracted evaluation question based on information that is sufficiently broad, deep, reliable, contextually relevant, culturally sensitive, and valid
		A1.2 Derive defensible conclusions that respond to the evaluation's stated purposes, e.g., to identify and assess the program's strengths and weaknesses, main effects and side effects, and worth and merit
		A1.3 Limit conclusions to the applicable time periods, contexts, purposes, and activities
		A1.4 Identify the persons who determined the evaluation's conclusions, e.g., the evaluator using the obtained information plus inputs from a broad range of stakeholders
		A1.5 Identify and report all important assumptions, the interpretive frameworks and values employed to derive the conclusions, and any appropriate caveats
		A1.6 Report plausible alternative explanations of the findings and

Section	Standard	Checkpoint
		explain why rival explanations were rejected
	A2. Valid Information	A2.1 Through communication with the full range of stakeholders develop a coherent, widely understood set of concepts and terms needed to assess and judge the program within its cultural context
		A2.2 Assure—through such means as systematic protocols, training, and calibration--that data collectors competently obtain the needed data
		A2.3 Document the methodological steps taken to protect validity during data selection, collection, storage, and analysis
		A2.4 Involve clients, sponsors, and other stakeholders sufficiently to ensure that the scope and depth of interpretations are aligned with their needs and widely understood
		A2.5 Investigate and report threats to validity, e.g., by examining and reporting on the merits of alternative explanations
		A2.6 Assess and report the comprehensiveness, quality, and clarity of the information provided by the procedures as a set in relation to the information needed to address the evaluation's purposes and questions
	A3. Reliable Information	A3.1 Determine, justify, and report the needed types of reliability—e.g., test-retest, findings from parallel groups, or ratings by multiple observers—and the acceptable levels of reliability
		A3.2 In the process of examining, strengthening, and reporting reliability, account for situations where assessments are or may be differentially reliable due to varying characteristics of persons and groups in the evaluation's context
		A3.3 Assure that the evaluation team includes or has access to expertise needed to investigate the applicable types of reliability
		A3.4 Describe the procedures used to achieve consistency
		A3.5 Provide appropriate reliability estimates for key information summaries, including descriptions of programs, program

Section	Standard	Checkpoint
		components, contexts, and outcomes
		A3.6 Examine and discuss the consistency of scoring, categorization, and coding and between different sets of information, e.g., assessments by different observers
	A4. Explicit Program and Context Descriptions	A4.1 Describe all important aspects of the program—e.g., goals, design, intended and actual recipients, components and subcomponents, staff and resources, procedures, and activities—and how these evolved over time
		A4.2 Describe how people in the program's general area experienced and perceived the program's existence, importance, and quality
		A4.3 Identify any model or theory that program staff invoked to structure and carry out the program
		A4.4 Define, analyze, and characterize contextual influences that appeared to significantly influence the program and that might be of interest to potential adopters, including the context's technical, social, political, organizational, and economic features
		A4.5 Identify any other programs, projects, or factors in the context that may affect the evaluated program's operations and accomplishments
		A4.6 As appropriate, report how the program's context is similar to or different from contexts where the program is expected to or reasonably might be adopted
	A5. Information Management	A5.1 Select information sources and procedures that are most likely to meet the evaluation's needs for accuracy and be respected by the evaluation's client group
		A5.2 Ensure that the collection of information is systematic, replicable, adequately free of mistakes, and well documented
		A5.3 Establish and implement protocols for quality control of the collection, validation, storage, and retrieval of evaluation information

Section	Standard	Checkpoint
		A5.4 Document and maintain both the original and processed versions of obtained information
		A5.5 Retain the original and analyzed forms of information as long as authorized users need it
		A5.6 Store the evaluative information in ways that prevent direct and indirect alterations, distortions, destruction, or decay
	A6. Sound Designs and Analyses	A6.1 Create or select a logical framework that provides a sound basis for studying the subject program, answering the evaluation's questions, and judging the program and its components
		A6.2 Plan to access pertinent information sources and to collect a sufficient breadth and depth of relevant, high quality quantitative and qualitative information in order to answer the evaluation's questions and judge the program's value
		A6.3 Delineate the many specific details required to collect, analyze, and report the needed information
		A6.4 Develop specific plans for analyzing obtained information, including clarifying needed assumptions, checking and correcting data and information, aggregating data, and checking for statistical significance of observed changes or differences in program recipients' performance
		A6.5 Buttress the conceptual framework and technical evaluation design with concrete plans for staffing, funding, scheduling, documenting, and metaevaluating the evaluation work
		A6.6 Plan specific procedures to avert and check for threats to reaching defensible conclusions, including analysis of factors of contextual complexity, examination of the sufficiency and validity of obtained information, checking on the plausibility of assumptions underlying the evaluation design, and assessment of the plausibility of alternative interpretations and conclusions
	A7. Explicit	A7.1 Clearly describe all the assumptions, criteria, and evidence that provided the basis for judgments and conclusions

Section	Standard	Checkpoint
	Evaluation Reasoning	A7.2 In making reasoning explicit, begin with the most important questions, then, as feasible, address all other key questions, e.g., those related to description, improvement, causal attributions, accountability, and costs related to effectiveness or benefits
		A7.3 Document the evaluation's chain of reasoning, including the values invoked so that stakeholders who might embrace different values can assess the evaluation's judgments and conclusions
		A7.4 Examine and report how the evaluation's judgments and conclusions are or are not consistent with the possibly varying value orientations and positions of different stakeholders
		A7.5 Identify, evaluate, and report the relative defensibility of alternative conclusions that might have been reached based on the obtained evidence
		A7.6 Assess and acknowledge limitations of the reasoning that led to the evaluation's judgments and conclusions
	A8. Commun- icating and Reporting	A8.1 Reach a formal agreement that the evaluator will retain editorial authority over reports
		A8.2 Reach a formal agreement defining right-to-know audiences and guaranteeing appropriate levels of openness and transparency in releasing and disseminating evaluation findings
		A8.3 Schedule formal and informal reporting in consideration of user needs, including follow-up assistance for applying findings
		A8.4 Employ multiple reporting mechanisms, e.g., slides, dramatizations, photographs, PowerPoint©, focus groups, printed reports, oral presentations, telephone conversations, and memos
		A8.5 Provide safeguards, such as stakeholder reviews of draft reports and translations into language of users, to assure that formal evaluation reports are correct, relevant, and understood by representatives of all segments of the evaluation's audience
		A8.6 Consistently check and correct draft reports to assure they are impartial, objective, free from bias, responsive to contracted evaluation questions, accurate, free of ambiguity, understood by

Section	Standard	Checkpoint
		key stakeholders, and edited for clarity
Evaluator Account-ability	E1. Evaluation Document-ation	E1.1 Document and preserve for inspection the following: Contract or memorandum of agreement that governed the evaluation
		E1.2 Evaluation plan, including evaluation tools and resumes of key evaluation staff
		E1.3 Evaluation budget and cost records
		E1.4 Reports, including interim and final reports, the evaluation's internal metaevaluation report, and, if obtained, a copy of the external metaevaluation report
		E1.5 Other information determined to be needed by reviewers, such as technical data on the employed evaluation tools, a glossary of pertinent theoretical and operational definitions involved in the evaluation, a description of the subject program, a record of stakeholder involvement, and news accounts related to the evaluation
		E1.6 Evidence of the evaluation's consequences, including stakeholders' uses of findings
	E2. Internal Metaeval-uation	E2.1 At the evaluation's beginning, determine the metaevaluation's intended users and uses, e.g., formative and summative
		E2.2 Develop a plan for obtaining, processing, and reporting a sufficient scope and depth of information to assess the evaluation's utility, feasibility, propriety, and accuracy and address the intended users' needs for timely metaevaluation feedback and reports
		E2.3 Assign responsibility for documenting and assessing the evaluation's plans, process, findings, and impacts and budget sufficient resources to carry out the internal metaevaluation
		E2.4 Maintain and make available for inspection a record of all internal metaevaluation steps, information, analyses, costs, and

Section	Standard	Checkpoint
		observed uses of the metaevaluation findings
		E2.5 Reach, justify, and report judgments of the evaluation's adherence to all of the metaevaluation
		E2.6 Make the internal metaevaluation findings available to all authorized users
	E3. External Metaeval-uation	E3.1 Confirm through exchange with key stakeholders the need for an external assessment of the evaluation and the purposes it should serve e.g., formative or summative
		E3.2 Stipulate that these and possibly additional standards will be used to assess and judge the evaluation
		E3.3 Select, recruit, and reach a formal agreement with an external metaevaluator who possesses an independent perspective, appropriate expertise, and freedom from possibly compromising connections or interests
		E3.4 Assure that the external metaevaluation is adequately planned, staffed, and funded
		E3.5 Provide the external metaevaluator with access to information and personnel required to conduct a thorough, defensible metaevaluation that serves the intended purposes
		E3.6 Assure that the metaevaluation will be subjected to appropriate quality control and that the metaevaluator will deliver as part of the metaevaluation report an attestation of its adherence to the metaevaluation standards

Stufflebeam (2011). Used with permission.

Potential Limitations

There are potential limitations to this research project. The most significant limitation is that the metaevaluation was conducted using a single case study, rather than on multiple evaluations conducted in a variety of settings. This limits the generalizability of the research findings. This limitation is somewhat unavoidable because of the relative newness of the model

and the limited amount of empirical research that has been conducted with it. However, researchers are increasingly making the case that single-case study research is valuable and worthwhile. Wong (2010) states that results of single case studies "might not be as conclusive as their more elaborate counterparts" but that they "still represent a significant advance for the field by making it easier for practitioners to evaluate their own practice and thereby encouraging more to do so" (p. 249).

A second limitation is the inherent potential for bias that exists in the metaevaluation process. As the evaluator, I have some level of control over the information that is delivered to the professional metaevaluator and to the evaluation consumers. Therefore, I have the ability to affect and be affected by the outcome of the evaluation, and hence become one of the stakeholders in the process. This is ultimately unavoidable in metaevaluation. However, since the purpose of metaevaluation is to determine "whether, on balance, after summarizing judgments across scales, the evaluation seems to achieve its purposes at an acceptable level of quality" (Fitzpatrick et al., 2011, p. 373), this is a limitation that is necessarily tolerated within the field.

A third limitation of the research project is the relative inexperience of the researcher. Brinkerhoff, Brethower, Hluchyj, and Nowakowski (1983) state that "not only should (metaevaluators) be competent enough to do the original evaluation, but they also have to be able to tell if it was a good or bad one and be able to convince others that they know the difference" (p. 208). I have conducted several evaluation projects during the course of graduate studies, but they do not qualify me as an expert in the field. In order to mitigate this limitation, I hired a professional metaevaluator to review the findings and participate in the study.

Definition of Terms

One-to-One (1:1) Technology Program – The fundamental characteristic of a one-to-one, or "ubiquitous", technology program is that students and teachers have their own internet-connected wireless computers in the classroom as well as access to a computer 24 hours a day, seven days a week. It is assumed that the students are using some sort of portable laptop computer that is loaded with current software. There is a focus in these programs on using the computers to help students complete academic tasks such as homework assignments, lab projects, test-taking, and presentations (Abell Foundation, 2008).

Evaluation Theory- Almost all evaluation theories are prescriptive, that is, they offer guidance, rules, and advice about what should and should not be done as part of an evaluation. None of the approaches are predictive or offer a true empirical theory. However, it is the convention of scholars in the field of evaluation to refer to their prescriptive approaches as theories (Christie & Alkin, 2008). In this study, I used the terms theory, model, approach, and design interchangeably.

Impact Evaluation – The purpose of an impact evaluation is to assess the changes that can be attributed to a certain program. This type of evaluation stands in contrast to outcome monitoring, which simply examines whether targets have been reached (Gertler, Martinez, Premand, Rawlings & Vermeersch, 2010). Impact evaluations not only look at immediate outcomes, but also at long-term program outcomes and the interdependence between the two (Guerra-Lopez, 2008).

Metaevaluation – A metaevaluation is a systematic review of an evaluation for the purpose of determining the quality of the processes and findings of the evaluation (Cooksy & Caracelli, 2009). When used as part of a single study, metaevaluations can serve either a formative or summative purpose. Formative metaevaluations are designed to improve the evaluation while it

is in process. Summative metaevaluation of a single study "provides information about the strengths and weaknesses of the evaluation to evaluation clients and audiences and to the evaluator" (Cooksy & Caracelli, 2009, p. 2).

Summary

This study is intended to address an ongoing challenge within the field of performance improvement; that is, to explicitly connect research and practice by basing practice in the field solidly on continuing empirical research. A new model within the field has been proposed, but needs additional research to confirm its effectiveness and identify its position within the field. Research questions were developed which allowed the researcher to test the model's utility, feasibility, propriety, accuracy, and accountability. The theoretical foundation for the study is based on the process of metaevaluation. Key terms that will be used in the context of the study have been defined. A review of relevant literature follows in the next section.

CHAPTER 2 Review of the Literature

This chapter will examine the literature that is relevant to this study. The literature review includes four sections. The first section is a review of evaluation theory. The goal of this review is to place the Guerra-Lopez Impact Evaluation Process within the body of evaluation theory literature. The second section examines the empirical evaluation of evaluation theory and models. The third section describes research about metaevaluation. Finally, since the first phase of the research study includes an evaluation of a 1:1 technology program, I will review the types of research that have been conducted to date on these programs.

Evaluation Theory

As evaluation began to emerge as a field of study in the mid twentieth century, it is not surprising that scholars disagreed about ideology and definitions. For example, from 1960-1990 more than fifty evaluation models were proposed (Fitzpatrick et al., 2011). This proliferation of models caused confusion among practitioners who were charged with determining which model was best for their purposes. The models developed based on the worldviews of their authors, and represented diverse philosophical orientations and methodological preferences. These differences have led scholars to propose a wide variety of designs, data collection methods, analysis methods, and techniques for interpretation. Through many years of scholarly debate, certain approaches have emerged as most commonly used and researched. However, the core differences remain and can be categorized along some common lines.

A variety of approaches to determining worth and merit are at the root of the diversity of views about evaluation. Fundamental beliefs about epistemology affect how individuals approach evaluation. Objectivists are interested in finding the truth and are drawn to methods that are scientifically objective and that yield results that are reproducible and verifiable.

Constructivists, on the other hand, believe that meaning is either developed individually or negotiated by a group, and are more interested in understanding the model in relation to its perceived effectiveness by its users.

A distinction that is closely related to the objectivist/constructivist debate concerns principles for assigning values. A utilitarian approach ascribes value by measuring the overall impact of a program. This approach tends to follow the objectivist epistemology. House (1976) suggests that utilitarian evaluation accepts the premise that the best option is the option that will benefit the most people. An intuitionist-pluralist approach, on the other hand, holds that value depends on the impact that the program has on each individual. This approach tends to follow the constructivist epistemology. Fitzpatrick et al. (2011) explain that these differences in philosophy caused rifts among scholars that affected the field for decades.

Differences in philosophy about knowledge and value will lead directly to differences in the choice of evaluation methods. The debate between quantitative and qualitative methods raged for many years among evaluation scholars. Mark (2001) claims that even if it hasn't been the *most* discussed topic in the field, it has certainly been the *loudest* discussion in the field. In recent years, however, it appears that the polarization has given way to an integration of approaches (Fitzpatrick et al., 2011). Scriven (1991) points out that evaluation is not a traditional discipline, but a "transdiscipline" that cuts across traditional lines, and that evaluators don't have the luxury of limiting themselves to one single inquiry paradigm.

An additional influence on the development of evaluation approaches has been the way that evaluators respond to the different needs that they perceive among their clients. Each audience has unique environmental contexts and stakeholder interests. The evaluator must

become familiar with the context and adapt the evaluation approach to it in order to successfully meet the client's needs.

Not surprisingly, these philosophical and methodological differences have led to a wide range of frameworks for classifying evaluation theories. More than a dozen evaluation theory classification schemata have been published, including: Guba and Lincoln (1981), House (1983), Scriven (1993), Alkin & Christie (2004), Christie & Alkin (2008, 2012), Stufflebeam & Coryn (2012), and Fitzpatrick et al. (2011). While each of these schemata provides interesting insight into evaluation theory, only three of them will be discussed in detail here: Fitzpatrick et al. (2011), Stufflebeam & Coryn (2012), and Christie & Alkin (2012). I chose to discuss Fitzpatrick et al. (2011) because it most clearly articulates the theoretical and methodological orientation of each category's proponents. The Stufflebeam & Coryn (2012) schemata is discussed because the metaevaluation that was conducted as part of this research study used the Stufflebeam methodology, and it is therefore worthwhile to understand his preferred classification schemata. Finally, the Christie & Alkin Evaluation Theory Tree (2012) is presented because it is intuitive, easy to understand, and provides a good visual representation of the major evaluation theories. Since the researcher's objective is to identify where the Guerra-Lopez IEP fits within the body of evaluation theory, the Christie & Alkin (2012) framework is particularly helpful. A side-by-side comparison of each of these frameworks and the scholars associated with them is included in Appendix B.

Fitzpatrick, Sanders, and Worthen Classification.

Fitzpatrick et al. (2011) divide the different approaches to evaluation into four categories:

- Program-oriented approaches - the focus is on identifying the purpose of a program, and then evaluating the extent to which the purpose is achieved

- Decision-oriented approaches – the primary goal is to meet the needs of managerial decision makers

- Consumer and Expertise-oriented approaches – In consumer-oriented evaluations, the focus is on providing evaluative information on "products" in order to help consumers make purchasing choices among competing products and services. In expertise-oriented evaluations, the focus is on relying on subjective professional expertise as the primary evaluation strategy

- Participant-oriented approaches – the program participants, or stakeholders, are primarily responsible for determining the evaluation criteria and goals

The program-oriented approach largely developed from the work of Ralph Tyler (1942). Tyler believed that evaluation was the process of determining the extent to which the predetermined objectives of a program had been obtained. His approach calls for establishing broad program objectives and then defining the objectives in terms of the behaviors that would indicate that the objectives have been reached. Next, his approach recommends selecting measurement techniques, collecting performance data, and finally comparing performance data with the behavioral objectives. Discrepancies between actual performance and behavioral objectives lead to program modifications. Malcolm Provus (1971) developed an alternative objectives-oriented approach that is referred to as the Discrepancy Evaluation Model. For Provus, evaluation is a process of agreeing on standards and then determining whether a discrepancy exists between the performance of an element of a program and the standards that were pre-set for the performance. This information is then used to decide whether to maintain, improve, or terminate the entire program or certain aspects of it.

The decision-oriented approach to evaluation holds that the purpose of evaluations should be to provide decision makers with information that will help them make sound decisions. This approach is largely based on systems theory, in that decisions are made about inputs, processes, and outputs. The most influential decision-oriented approach is Stufflebeam's (1971) CIPP model. Stufflebeam believes that evaluation is designed to provide useful information for judging decision alternatives. His evaluation framework is designed to help decision makers facing four kinds of decisions: planning, structuring, implementing, and recycling. The CIPP model focuses on four distinct but related activities, the evaluation of context, inputs, processes, and products. Evaluations of these activities can be done collectively, individually, sequentially, or simultaneously, depending on the stakeholder's needs.

The consumer-oriented approach is described by Fitzpatrick et al. (2011) as predominately a summative evaluation method since its purpose is to provide consumers with information that will help them choose among available alternative products. Scriven's early (1967) work distinguishing between formative and summative evaluation is the foundation of this approach. This approach is used extensively by government agencies and consumer advocate groups. The expertise-oriented approach depends on input provided by subject matter experts. Although all of the evaluation approaches discussed here rely on professional judgment to some extent, this approach is unique because it uses professional expertise as the primary evaluation strategy. This approach frequently calls for the use of a team of evaluators since it is unusual for one person to have all of the knowledge required to complete the evaluation adequately. Eisner's Educational Connoisseurship (1976) model falls into this category. He makes the case that an evaluator must bring both connoisseurship and criticism to the evaluation process. Connoisseurship is the art of being aware of an object's qualities and the relationships among

these qualities. This perception of quality stems from a perceptual acuity that is based on previous experience. Criticism, on the other hand, is "the art of disclosing the qualities of events or objects that connoisseurship perceives" (Eisner, 1979, p. 197). The expertise of the evaluator is obviously critical in this model because the validity of the model depends on the evaluator's perception.

Participant-oriented approaches are focused on identifying and articulating the needs, values, and perspectives of program stakeholders in order to make judgments about the merit or worth of the program. Unlike the approaches discussed above, this approach starts by considering the interests of the stakeholders. The approach stresses first-hand experience with program activities and settings. It often uses inductive reasoning, gathers information from multiple data sources, and does not follow a standard plan. Several of the current popular evaluation theories fall into this category, including Stake's (1975) Responsive Evaluation model, Guba & Lincoln's (1981) Naturalistic Evaluation, Fetterman's (1994) Empowerment Evaluation, and Mertens' (1999) Emancipatory Evaluation. Fitzpatrick et al. (2011) also include Patton's (1986) Utilization Focused Evaluation in this category. Patton's model will be discussed in greater detail in the following section.

According to the classification schemata proposed by Fitzpatrick et al. (2011), the researcher has concluded that the Guerra-Lopez IEP falls into the decision-oriented approach. The decision-oriented approach is closely related to the systems approach, and its greatest strength is that it provides clear focus to the evaluation. It also stresses the importance of the utility of information. Although it is tempting to place the IEP in the participant-oriented approach based on the centrality of the stakeholder in the process, Guerra-Lopez (2007a) stresses that a foundational principle of the IEP is that the evaluand (the program being evaluated) should

always be considered a means to an end, where the end manifests itself in the results that the organization achieves. The evaluand must ultimately add value at the strategic level, but in the short term it helps the organization establish a chain of impact at the operational and tactical levels. The alignment that the approach demands helps to focus the management decision making process, and in fact, the identification of key management decisions is the basis for identifying what data should be collected, and the evaluation plan as a whole.

Stufflebeam Classification

Stufflebeam & Coryn (2012) identify four categories of evaluation approaches. They are:

- Questions and Methods – the focus is on a narrowly defined set of questions or on a specific methodology

- Improvement and Accountability – the focus is on the assessment of a program's merit or worth, and are usually objectivist

- Social Agenda and Advocacy – the focus is on ensuring that all segments of society have access to social and educational opportunities and services

- Eclectic – Utilization-focused evaluation, which has elements of the other categories, but primarily focuses on ensuring that the evaluation has an impact

The Questions and Methods approach to evaluation is really two different approaches that Stufflebeam combines to simplify his schemata. Questions-oriented evaluations are intended to answer specific questions, often using a wide variety of methods. Methods-oriented approaches tend to use a single, pre-specified method. Stufflebeam groups the two approaches together because they both tend to narrow the scope of evaluations. Questions-oriented approaches usually begin with narrowly defined questions, which might be taken from a program's

operational objectives, a set of accountability requirements, or an expert reviewer's preferred evaluation criteria. A methods-oriented approach might use a design for a controlled experiment or a particular standardized test as its starting point. Technical quality is always emphasized in methods-oriented approaches. Both approaches stress that it is better to answer a few questions well than to conduct a broad assessment of a program's merit and worth (Stufflebeam, 2001a). Since the approaches rarely fully assess merit and worth, Stufflebeam refers to them as "quasi-evaluation studies" (2001a, p. 17). While they have a legitimate purpose, Stufflebeam feels that they should not be uncritically equated with real evaluation.

The Improvement/Accountability-oriented approaches differ from Questions and Methods because they stress the need to fully assess merit and worth. They are usually expensive and comprehensive in their consideration of a full range of criteria and questions. Usually they use the assessed needs of the program's stakeholders as the foundational criteria for the evaluation. They look at the full range of technical and economic criteria in order to judge operations and program plans. They usually begin from an objectivist theoretical perspective and thus seek definitive answers based on an assumed underlying reality. They often use a variety of quantitative and qualitative methods in order to crosscheck the evaluation findings (Stufflebeam, 2001a).

The Social Agenda/Advocacy approaches are intended to make a difference in society through program evaluation. They seek to address the inequality of access to educational and social opportunities and services that exist in our society. They are heavily oriented towards the perspective of the stakeholders. They are based on a constructivist theoretical perspective and almost always rely on qualitative methods. They do not seek to find the "right" or "best" answer, but instead emphasize cultural pluralism, multiple realities, and moral relativity. Stufflebeam

expresses a concern that these approaches might fail to meet the standards of a sound evaluation because they concentrate so heavily on fulfilling a social mission. However, they are oriented towards fairness and equity, and they tend to employ strong procedures for involving a full range of stakeholders (Stufflebeam, 2001a).

In their most recent iteration of Stufflebeam's schemata, Stufflebeam & Coryn (2012) identify a fourth category of evaluation that they consider to be among the best for use in the twenty-first century. Identified as the "eclectic" approach to evaluation, Stufflebeam only identifies one model that falls into the category. Patton's (1986) Utilization-focused approach is geared towards ensuring that evaluations have an impact. The Utilization-focused approach is a process for making decisions in collaboration with a targeted group of stakeholders, selected from a broader set of stakeholders, in order to focus on the intended use of the evaluation. All aspects of these evaluations are chosen so that they will help users apply the evaluation's findings to their intended uses. This approach does not fall neatly into any of the previous approaches identified by Stufflebeam because it contains elements of each. It has the Social Agenda/Advocacy's emphasis on the democratic participation of representative stakeholders. However, it does not necessarily advocate a particular social agenda, so it doesn't quite fit in this category. The approach agrees with the Improvement/Accountability approach in that it promotes maximizing impact. However, it takes a more pragmatic approach, and will draw on any legitimate approach to guarantee that it helps the stakeholders reach their goals (Stufflebeam, 2001a).

Based on the schemata proposed by Stufflebeam, the researcher believes that the Guerra-Lopez IEP falls into the eclectic category alongside Patton's Utilization-focused approach. The two approaches share several characteristics: they are primarily concerned with impact, they rely

on the input of representative stakeholders to focus the evaluation, and they use mixed methods to achieve the desired results. The two approaches differ somewhat in their focus on the stakeholders. Using the IEP, the evaluator is focused on identifying appropriate stakeholders, and helping them to identify the desired results of the program at the strategic, tactical, and operational levels. The focus is really on helping the stakeholders identify the program's "ends". Ultimately this identification of ends supports implementation and continuous improvement efforts. In the Utilization approach, the evaluator "engages the client group to clarify why they need the evaluation, how they intend to apply its findings, how they think it should be conducted, and what types of reports should be provided" (Stufflebeam, 2001a, p. 77). Both approaches are focused on impact and on stakeholders, but the IEP stresses the impact on organizational goals, clients and society, while the Utilization approach focuses on making sure that the results of the evaluation are *used* by the stakeholders to fulfill their stated objectives. This subtle difference in purpose distinguishes the approaches.

Alkin and Christie Evaluation Theory Tree

Alkin & Christie (2004) developed a framework of prescriptive evaluation theories that are depicted graphically as a tree, and which include the following categories:

- Methods – the primary concern is with research methodology so that the results will maximize generalizability or knowledge construction

- Values – the central feature is on the process of placing value on the subject of the evaluation (the evaluand). This branch is split into objectivist and subjectivist perspectives

- Use – the main focus is on the way in which the results of the evaluation will be used, and by whom

The original Evaluation Theory Tree was published in 2004. Christie & Alkin updated it in 2008, and again in 2012. A graphic representation of the current Evaluation Tree appears in Figure 3.

Figure 3. The Evaluation Theory Tree
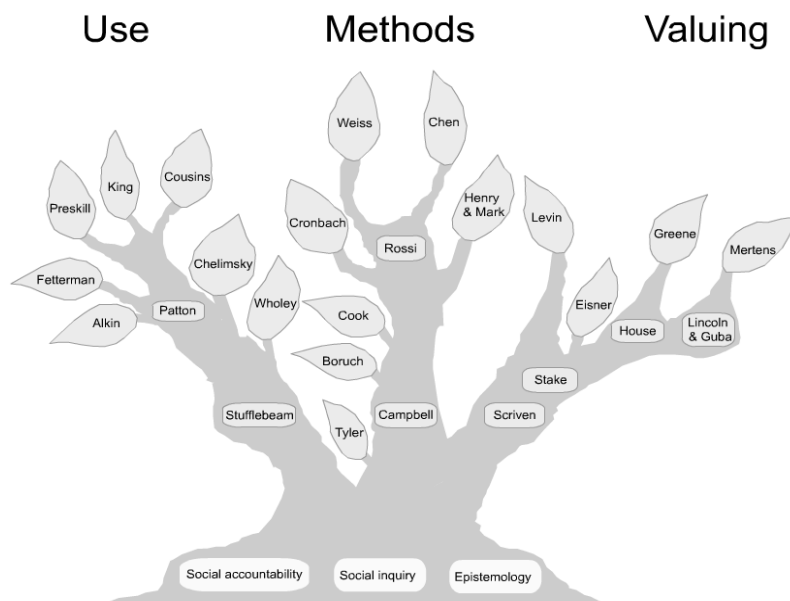
Christie, C. & Alkin, M. (2012). Used with permission.

Christie and Alkin (2012) use the metaphor of a tree to describe the purpose of evaluation theory. The graphic depicts the foundations of evaluation as social accountability, systematic social inquiry, and epistemology. Each of these represents an important reason to conduct evaluations, and has therefore supported the development of the field in its own way. Social accountability is an important motivation for evaluation, and it is a way to improve programs and society. Systematic social inquiry stems from a concern for using a justifiable set of procedures for determining accountability. Finally, all evaluation theorists must base their models on a basic

epistemology, which informs their purposes, assumptions, values, and methodology (Christie & Alkin, 2012).

The central branch of the tree is labeled the "Methods" branch, and it develops directly out of the social inquiry root. The theorists who are represented on this branch are primarily guided by research methodology. They are particularly concerned with obtaining the most rigorous results possible within the contextual constraints in order to contribute to knowledge construction. Most of the theorists represented on this branch work from an objectivist positivist epistemology. The foundational theorist on this branch is Donald Campbell, who is best known for his efforts to eliminate bias in the performance of research in field settings. The other theorists placed on this branch of the tree have a variety of perspectives, but are fundamentally concerned with using methods that ensure generalizability. As an interesting side note, Ralph Tyler is positioned on the tree as a small outgrowth of the methods branch. Christie and Alkin (2012) believe that his work was groundbreaking and essential to the development of the field, but that it did not necessarily influence the work of later theorists, and therefore deserves its own little branch, close to the foundations of the field.

The right hand branch of the tree is labeled "Values" and corresponds closely to Stufflebeam's categorization of Social Agenda and Advocacy approaches. Michael Scriven and Robert Stake are at the root of the branch, because of their work establishing the vital role of valuing in evaluation. The theorists on this branch are primarily concerned with placing value on the subject of the evaluation. The branch is split into objectivist and subjectivist sides. The objectivist side of the branch is closest to the methods branch, and is more concerned with identifying the merit and worth of the evaluand. The subjectivist side of the branch is also

concerned with measuring value, but takes a more relativistic viewpoint that recognizes that reality is a dynamic process and that truth is always relative to a frame of reference.

The "Use" branch of the tree is largely based on the work of Daniel Stufflebeam, and was originally oriented towards decision making. Fundamentally, the theorists represented on this branch are concerned with the way that the results of the evaluation will be used and particularly focuses on the people who will use the information. Recently, the concept of "use" in evaluation has expanded to include "evaluation influence", which refers to the capacity of the evaluation processes or findings to indirectly produce a change in the organization.

Based on the metaphor of the evaluation theory tree, the researcher has concluded that the Guerra-Lopez IEP would fit onto the Use branch, above Stufflebeam, near Patton's Utilization-focused approach. My conceptualization of the Evaluation Theory Tree with the IEP placed on it appears in Figure 4.

Figure 4. Evaluation Theory Tree Reconfigured to Include the Impact Evaluation Process
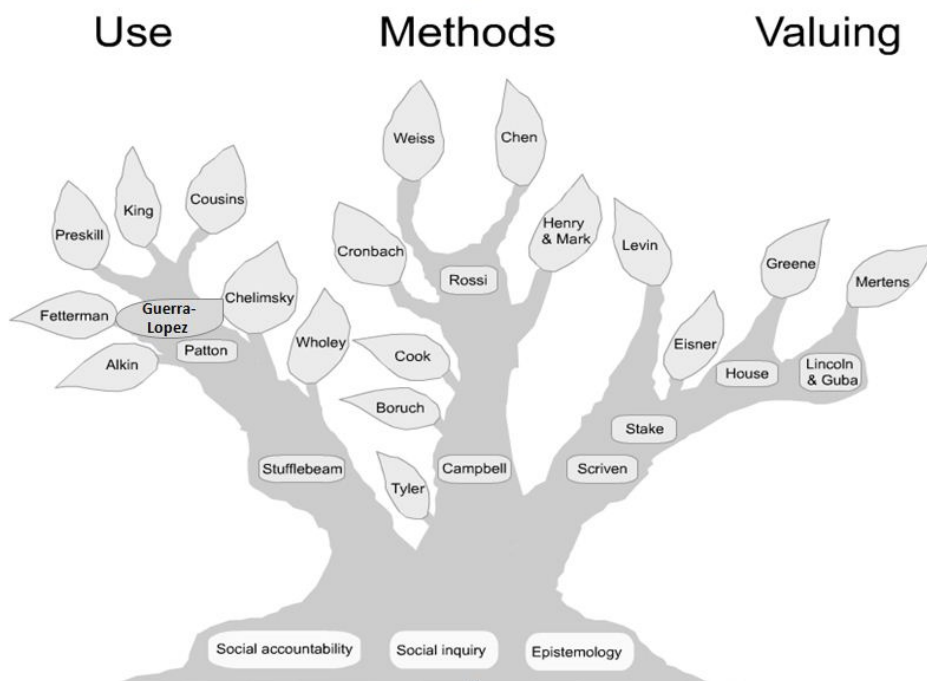
In summary, through this review of the evaluation theory literature, the researcher has been able to identify the appropriate position for the Guerra-Lopez Impact Evaluation Process within each of the theoretical frameworks that have been discussed. This analysis highlights the essential features of the model that include a fundamental concern with the impact that the evaluation will have on organizational results and the primary role of representative stakeholders in focusing the evaluation. The IEP is classified as a Decision-Oriented approach according to the Fitzpatrick et al. (2011) classification system. It is an Eclectic approach according to Stufflebeam & Coryn's (2012) framework. It is part of the Use branch of the Christie & Alkin (2012) schemata. Appendix B identifies the leading theorists, their corresponding theories, and their placement within each of the three classification schema examined.

The placement of a theory within a classification framework does not guarantee that it can make an effective transition from theory to practice. In order to assess a model's effectiveness, it is necessary to empirically evaluate it. Since the field of performance improvement is fundamentally concerned with solving problems related to human performance in the workplace (Mager, 2006), it is incumbent on scholars and practitioners within the field to use evaluation models that have been evaluated empirically.

Empirical evaluation of evaluation theory

The purposes of the current study include an examination of the theoretical underpinnings of the IEP as well as an evaluation of its effectiveness in practice. By addressing these dual purposes, the research study crosses the divide between theory and practice. The study of evaluation theories is valuable because it provides evaluators with ideological perspectives that

can inform the multitude of decisions that they have to make about how to design and conduct their own evaluations (Smith, 2007). Evaluation theories can also provide guidance about the appropriate role of the evaluator, and the relationship between the evaluator and the evaluand (Ryan & Schwandt, 2002). They can help evaluators select evaluation questions and pair these questions with appropriate methods (Greene, 2005; Henry, Julnes & Mark, 1998). The study of evaluation theory can help evaluators identify who should participate in determining the direction of the evaluation (Cousins & Whitmore, 1998, Fetterman, 1994), and how, when, and to whom the evaluation results be delivered (Patton, 2008).

It seems obvious that serious consideration of alternative theoretical perspectives should result in evaluations that differ distinctly on key dimensions and inform the practice of evaluation. Miller (2010) states that

> Sorting through theories and determining their ultimate feasibility and merit
> would benefit by close empirical examination of how evaluation theories can be
> and are applied in practice, whether they consistently and reliably lead to
> successful evaluation, and under what circumstances "good" evaluation are likely
> to emerge. (p. 391)

There is a reciprocal relationship between the study of theory and the study of practice. Understanding practice and the use of theory in practice can help to inform the development of future theories. Miller (2010) proposes a framework for empirically evaluating how evaluation theory informs practice and whether particular theories yield superior evaluations. Miller's framework consists of 5 criteria: operational specificity, range of application, feasibility in practice, discernible impact, and reproducibility.

Operational specificity refers to the concept that, in order for a theory to be useful in practice, it must offer specific guidance for practice. This includes providing procedural guidelines regarding how and when evaluations are conducted, what questions are used, how they are prioritized, who is expected to participate in the evaluation, what the role of the evaluator is, what methods will be used, how the values that underlie the theory are best enacted and how the evaluation will be used. Empirical evaluation of theory "requires precise articulation of the implications for practice inherent in the theory, as well as the identification of operational ambiguities" (Miller, 2010, p. 392).

Range of application addresses the fact that there is no single theory that is ideally suited to every application. The empirical evaluation of theory, therefore, must consider the limits of the theory's application. Miller suggests asking: What conditions are most suitable for applying this theory? How adaptable is it across a wide range of conditions? Will the theory have a different outcome when it is applied under ideal circumstances than when it is applied under those that are less than ideal?

Feasibility in practice involves including an assessment of how easy or difficult the prescriptions for practice are to apply in real life. Is it really possible for the evaluator to do what the theory requires? Miller points out that "the technical, ethical, skill, and resource requirements associated with these designs have implications for the evaluation circumstances under which it is feasible to follow theoretical prescriptions" (p. 394). For example, conducting a utilization-focused evaluation is easier to do if you possess political savvy, expertise, and outstanding interpersonal skills. However, if the evaluator lacks any of these qualities, the evaluation will become markedly more difficult.

Discernible impact is concerned with the close examination of whether the use of a theory really leads to the impact that is expected or whether unintended effects occur. Many theories emphasize the impact that occurs as a direct result of the way the evaluation is conducted. For example, House and Howe (1999) emphasize the goal of promoting democratic dialogue, Preskill & Torres (2001) seek to facilitate organizational learning, Mertens (2001) argues for the transformation of social arrangements, and Patton (1986) attempts to improve evaluation's influence. Theoretically speaking, there should be discernible benefits "because of, due to, and linked to" (Miller, 2010, p. 395) any evaluation approach that is implemented as intended by its developer.

Reproducibility refers to the important component of determining whether any impacts that are observed as a result of an evaluation can be reproduced over time, place, and evaluator. It is important to know what diverse evaluators actually do when they use an approach, as well as to know whether their implementation approximates the standards set for the approach, and whether the approach can achieve its intended outcomes in another evaluator's hands. Miller (2010) believes that close examination of reproducibility of theories "may help to categorize theories regarding the degree to which they are primarily useful as sensitizing ideologies or sources of practical guidance on carrying out aspects of evaluation" (p. 396).

Miller's (2010) call to empirically evaluate evaluation theories should appeal to performance improvement professionals. As part of the current evaluation, I addressed the elements of Miller's framework in relation to the IEP.

Metaevaluation

The term "metaevaluation" was coined by Michael Scriven (1967) in reference to an evaluation that he had done on a plan for evaluating educational products. He believed that it was possible for evaluators to issue inaccurate or biased reports that could lead consumers to purchase inferior educational products and use them to the detriment of young people in society. It is necessary, therefore, that evaluations of such products should themselves be evaluated (Stufflebeam, 2000).

The evaluation literature is full of advice about how to conduct a metaevaluation. Scriven (1991) distinguishes between formative and summative metaevaluations and recommends the use of standards against which to measure performance. The *Program Evaluation Standards* published by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 2010) is the most commonly referenced set of standards used in metaevaluation. Stufflebeam is undoubtedly considered to be the "father" of metaevaluation. Throughout his career he has made a strong plea for the increased use of metaevaluation to guide evaluations and to report their strengths and weaknesses. He also developed a structure of metaevaluation procedures which are based on the Joint Committee's standards.

Nilsson and Hogben (1983) make a crucial distinction that is central to this research study. They point out that everyone refers to metaevaluation as "the evaluation of evaluation" but that they miss the fact that "the evaluation of evaluation can mean both the evaluation of evaluation itself and the evaluation of particular specimens of evaluation". While this statement seems confusing at first glance, it is at the heart of the current study. A metaevaluation can look at a single evaluation of a program or product, or it can look at an evaluation model itself. While

the current study does both, the primary purpose is the latter. The distinction first articulated by Nilsson and Hogben (1983) has resurfaced in more recent articles and books. Mark and Henry (2004) make the case that there needs to be more research on mechanisms of evaluation influence. Metaevaluation is an excellent way to gather the data that would help to identify these mechanisms that will ensure that evaluations achieve their intended purpose.

In spite of the frequent call for metaevaluation in the literature (Cooksy & Caracelli, 2009; Fitzpatrick et al., 2004; Nilsson & Hogben, 1983; Stufflebeam, 2001b) and the broad possibilities for expanding the use of metaevaluation (Mark & Henry, 2004), there are few actual examples of metaevaluations in the evaluation literature. In 1999, the *American Journal of Evaluation* introduced a new section titled "Evaluating Evaluations" in response to the lack of published metaevaluations. The section was designed to demonstrate the benefits and utility of metaevaluations and to improve evaluation practice. It presented articles that summarized metaevaluation efforts (Cooksy, 1999; Grasso, 1999), but was short-lived.

Stufflebeam (2000) describes seven metaevaluations that he conducted, including the teacher evaluation system employed by Teach For America, the United States Marine Corp's officer and enlisted personnel performance evaluation system, the Hawaii Department of Education's system to evaluate public school teachers, and Australia's evaluation of a national distance baccalaureate program called Open Learning Australia, among others. He used the lessons learned from these metaevaluations to develop a list of Ten Main Steps in metaevaluation, which are shown below in Table 4. Here again, the point is made that conducting metaevaluations can improve both particular evaluations and the practice of evaluation in general.

Table 4. Stufflebeam's Ten Main Steps in Metaevaluation

Step	Activity
1	Determine and arrange to interact with the metaevaluation's stakeholders
2	Establish a qualified metaevaluation team
3	Define the metaevaluation questions
4	Agree on standards to judge the evaluation system or particular evaluation
5	Frame the metaevaluation contract
6	Collect and review pertinent available information
7	Collect new information as needed, including, for example, on-site interviews, observations, and survey
8	Analyze the qualitative and quantitative information and judge the evaluation's adherence to the selected evaluation standards
9	Prepare and submit the needed reports
10	Help the client and other stakeholders interpret and apply the findings

Lynch et al. (2003) conducted a descriptive case study metaevaluation of an interdisciplinary curriculum for a rural health training program. The authors used the *Program Evaluation Standards* (Yarbrough et al., 2011) as a framework for the metaevaluation, and found that the results illustrated the breadth and interrelatedness of issues involved in curriculum evaluation. The stakeholders indicated that they would use the results of the metaevaluation as a starting point for further improvements in their regular evaluation process.

A final example of metaevaluation that appears in the literature is the Advanced Technological Education (ATE) Evaluation project, which paid for four external metaevaluations of its projects (Gullickson, Wingate, Lawrenz, & Coryn, 2006). The emphasis in these studies was on the validation of processes used. The project considered what had been done as part of the original evaluation and judged the quality of individual techniques rather than focusing on how the overall evaluation might be improved.

The literature is littered with studies that claim to be metaevaluations but are more like meta-analyses (Ashworth, Cebulla, Greenberg & Walker, 2004; Brandon, 1998; Woodside & Sakai, 2001). Metaevaluation assesses an evaluation's merit and worth, while a meta-analysis is a synthesis of studies that address a common research question (Stufflebeam, 2001b).

Since the current study is, in effect, the evaluation of an evaluation model, it is crucial to look at the literature to see what has been done in this regard. Fitzpatrick et al. (2011) propose five steps for conducting a metaevaluation of an evaluation design. First, the evaluator needs to obtain a copy of the evaluation design in a form that is easy to review. In the next two steps, evaluators need to identify who will conduct the metaevaluation and make sure that they have authorization to evaluate the design. Next, the evaluator selects the standards to use in the metaevaluation. Fitzpatrick et al. (2011) suggest that the following *Program Evaluation Standards (*Yarbrough et al., 2011) are most relevant for evaluating designs:

- A6- Sound Design and Analyses,
- U3 – Negotiated Purpose,
- P1 – Responsiveness and Inclusion Orientation,
- P3 – Human Rights and Respect,
- P4 – Clarity and Fairness
- P5 – Conflicts of Interest
- A1 – Justified Conclusions and Decisions
- A2 – Valid Information
- A3 – Reliable Information
- A4 – Explicit Program and Context Descriptions
- A5 – Explicit Evaluation Reasoning
- U2 – Attention to Stakeholders
- U4 – Explicit Values
- U6 – Meaningful Processes and Purposes
- F1 – Project Management
- F2 – Practical Procedures
- F3 – Contextual Viability
- F4 – Resource Use

Finally, based on review of all of the above, the evaluator judges the adequacy of the design. Since no design is perfect, the purpose is to determine "whether, on balance, after summarizing judgments across scales, the evaluation seems to achieve its purposes at an acceptable level of quality" (Fitzpatrick et al., 2011, p. 373).

Stufflebeam is the pre-eminent proponent of evaluating evaluation models. Stufflebeam and Coryn (2012) conducted a comparative analysis and evaluation of nine evaluation approaches that they deemed to be the best approaches for 21st century evaluations. Their choices for best evaluation model included the success case method, case study, experimental and quasi-experimental design, objectives-based, the CIPP model, consumer-oriented evaluation, constructivist and responsive/client centered approach, and utilization focused evaluation. Each of these approaches is described in more detail in Appendix B. The goal of their analysis was to help evaluators and their clients critically appraise each of the approaches before choosing among them. Stufflebeam and Coryn's (2012) method for evaluating the models included the use of Stufflebeam's (2011) *Checklist*, which is keyed to the revised *Program Evaluation Standards* (Yarbrough et al., 2011). The authors independently rated each of the nine approaches against the 30 standards based on their experience "seeing and assessing how these approaches worked in practice" (Stufflebeam & Coryn, 2012, p. 443). Subsequently, they jointly reviewed their ratings, reached consensus on discrepancies, and finally calculated a numerical rating for each approach. The use of Stufflebeam's (2011) newly revised *Checklist* is one of the key elements of the current study.

Newman (2011) recognized the need to "ensure the efficacy of evaluation models to appropriately assess the programs which they are intended to evaluate" (p. vi). His research was designed to empirically test the theory behind Guskey's (2003) professional development model,

which is commonly used by school systems to assess programs. The study focused on testing the nomological network of the Guskey model. A nomological network is a representation of the key concepts in a study, their observable manifestations, and the relationships among these concepts and manifestations. Newman investigated the relationships among five key components of Guskey's Model (Teacher Satisfaction, Teacher Knowledge, Teacher Practices, Administrative Support and Student Outcomes) and found strong support for the model's continued use.

In summary, while there is widespread support for the concept of metaevaluation, there is a dearth of actual research that uses metaevaluative methods. In addition, while researchers call for further metaevaluation of evaluation models, there is almost no current research along these lines. This is not because the topic is not important or valuable, but seems to be more related to the fact that metaevaluations are expensive and complex, and as a result they are more likely to be done on individual evaluations than on models.

1:1 Technology Programs and Ubiquitous Computing

Since one of the purposes of this research study is to assess the effectiveness of the Impact Evaluation Process as a tool to evaluate a one-to-one (1:1) technology program, it is important to review the literature concerning the evaluation of these programs. 1:1 technology programs are a relatively new phenomenon, and the first published evaluation of such a program occurred in 1998 (Rockman, 1998). The study evaluated the results of a three year partnership between Microsoft Corporation and Toshiba America Information Systems. They sponsored a pilot project to provide laptops to students in a variety of locations across the United States. The first year evaluation simply looked at implementation. The second and third year evaluations began to examine impacts on teaching and learning as well as whether or not the laptops were supporting a constructivist pedagogy. The evaluation found that students became more confident

in their use of technology, but results about the impact of the technology program on standardized test scores were inconclusive (Blumenthal, 2003). Many of the early laptop programs were sponsored by computer companies, such as Apple Corporation, and some of the early evaluations were conducted by consulting groups, presumably hired by the sponsors (Blumenthal, 2003). This calls into question the results of these early evaluations.

More recently, research on 1:1 technology programs has largely focused on individual factors, rather than on a comprehensive examination of overall effect. For example, McLaren (2011) examined the effect of a technology program on student interaction in institutions of higher education in the United Arab Emirates. Skevakis (2010) isolated teacher perceptions of principal leadership behaviors that were associated with the integration of a technology program. Dalgarno (2009) and Pogany (2009) looked at teacher response to laptop initiatives.

One of the most frequently stated goals of 1:1 technology programs is to improve students' "21st Century skills". There is no universal agreement about what constitutes a 21st Century skill, but several researchers and organizations have identified some possibilities. The Abell Foundation (2008) includes the following in their list of necessary skills for success in the 21st century: problem solving skills, the ability to research information independently, the ability to use resources related to real-life skills, the ability to utilize technology. The Partnership for 21st Century Skills (2012) lists the following: critical thinking and problem solving, creativity and innovation, communication, collaboration, information literacy, media literacy, flexibility and adaptability, initiative and self-direction, social and cross cultural interaction, productivity and accountability, and leadership and responsibility. Several recent research studies have attempted to measure the effect of a 1:1 technology program on these skills, including Staib

(2011) and Chandrasekhar (2009). These studies are both phenomenological case studies, which ultimately focused on student and teacher engagement.

Another frequently cited reason for implementing a 1:1 technology program is to increase student achievement. Some of the studies that examine achievement focus on one or two subjects in particular. Queener (2011) looked at student achievement in mathematics, and found that achievement scores were affected initially, but the changes were not sustained over time. Holcomb (2009) looked at math and writing test scores and found that the technology programs had a positive effect on both. Mills (2010), on the other hand, took a more global look at how a laptop initiative affected student achievement and academic performance. His study combined standardized test scores and the students' own perceptions about their achievement level, which provided a triangulated view of achievement. Harris (2010) approached the achievement question from a social justice perspective and examined achievement levels at several points along the "digital divide", which separates affluent students from low socioeconomic students. He found that technology programs affected the achievement levels of low socioeconomic students more than they did the achievement levels of their more affluent counterparts.

Since one of the main purposes of the current research project is to study a specific evaluation model, the researcher is more interested in the technology program literature that focuses on *how* the program evaluations were conducted than on those that focused on elements of the programs. It is immediately apparent that evaluations that focus on student achievement levels tend to be more quantitative, while evaluations that examine perceptions are more likely to be qualitative in nature. However, beyond this basic observation, it is difficult to discern a research framework that is widely used to evaluate technology programs.

In response to this lack of a common framework for technology program evaluations, the Rand Corporation funded a Digital School District initiative in Pennsylvania in 2001-2002 (Kerr, Pane & Barney, 2003). During the second year of implementation, Rand produced a summative evaluation of the program, as well as a research design for future evaluation. Rand researchers developed a conceptual framework for further evaluation of the initiative, which they called a Theory of Change, based on the proposal that was developed by their partner school district and on the anticipated outcomes of the program. The Theory of Change was the basis for a set of research questions and developed into a table of possible data sources that could be used to measure the effects of the initiative. Unfortunately, there is no evidence that the Rand framework was used again in further published studies.

A second framework was proposed by Zucker (2004). It appears in Figure 5 below. It is based in part on discussions that the author had as part of a National Science Foundation sponsored project with a consortium of researchers from multiple institutions. The framework is intended to be used at the system or school district level rather than at the classroom level.

Figure 5. Framework for Research and Evaluation of One-to-One Computing

A Framework for Research and Evaluation of One-to-One Computing
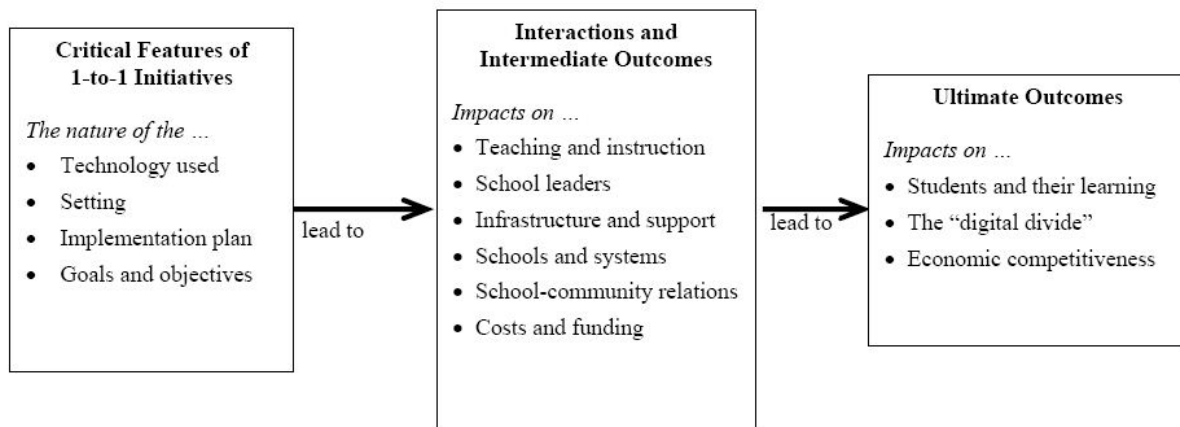
The box on the left of Figure 5 represents the features that are critical for 1:1 programs. Since not all 1:1 programs are the same, it is important to understand the critical features that distinguish among them. For example, if the 1:1 program is based on handheld devices, it will have a different impact than a program based on laptops. Other critical features include the setting, the plan that the school or district developed to implement the program, and the goals and objectives that are identified at the outset of the program.

The box on the far right represents the most important expected goals of the 1:1 computing program. These include the students and their learning, which can take a variety of forms, and include changes in test scores, acquisition of 21st Century skills, and increased student motivation. Another ultimate outcome is to narrow the gap in the digital divide, and is based on increasing equity of access to computing and information. Economic competitiveness is the final ultimate outcome. The framework intentionally limits the ultimate outcomes to a small number to signal that research that focuses on these goals is especially valuable (Zucker, 2004).

The middle box in Figure 5 identifies the intermediate outcomes of a technology program and, according to Zucker (2004), answers the question *how*. While these intermediate outcomes may be considered desirable in their own right, they are ultimately means that will lead to the ultimate ends, or outcomes, listed on the right.

The Zucker (2004) framework shares some similarities with the Guerra-Lopez IEP. Both researchers make the critical distinction between means and ends, and highlight the primary goal of evaluating whether or not the ends have been achieved. Both identify ends at multiple levels: strategic, tactical, and operational. For both researchers, strategic goals involve ensuring societal benefits. For Zucker, this includes assuring access to technology for low socioeconomic students.

For Guerra-Lopez this corresponds to Kaufman's ideal vision of "the world we want to create for tomorrow's child" (Guerra-Lopez, 2008, pp. 21-22).

In summary, while there is a substantial amount of literature on 1:1 technology programs, it is largely aimed at examining specific elements of the programs rather than examining overall impact, and is rarely based on an identified conceptual framework. The current study will add to the research base that uses a clearly identified conceptual framework to conduct 1:1 technology program evaluations. The following chapter describes the methodology that was used to conduct the study.

CHAPTER 3 Methodology

Introduction

The purpose of this study is to assess the effectiveness of the Guerra-Lopez Impact Evaluation Process (IEP) as a tool to evaluate programs, and to conduct a metaevaluation of the IEP based on the 2011 Joint Committee's *Program Evaluation Standards*.

The research questions that follow from this statement of purpose are:

Research Question 1: Where does the Guerra-Lopez Impact Evaluation Process (IEP) fit into the body of evaluation research?

Research Question 2: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Utility?

Research Question 3: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Feasibility?

Research Question 4: To what extent does the IEP meet the Joint Committee Program Evaluation requirements for Propriety?

Research Question 5: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accuracy?

Research Question 6: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accountability?

Research Question 7: How effective is the IEP for evaluating educational programs?

Research Question 1 was addressed as part of the literature review. The remaining research questions were addressed in the current study. The study consisted of two phases. During the first phase, I conducted an evaluation of a 1:1 technology program at a private secondary school in suburban Detroit using the Guerra-Lopez Impact Evaluation Process (IEP). I

engaged a group of more than 30 representative stakeholders to participate in the evaluation. During the second phase, I empirically evaluated the IEP by conducting a metaevaluation of the evaluation design using Stufflebeam's (2011) newly revised *Program Evaluation Metaevaluation Checklist.*

Chapter 3 describes the methodology of this research study, and includes descriptions of: (a) the rationale for the study's qualitative design, (b) the rationale for the single case study design, (c) a description of the research setting, (d) a description of the research participants, (e) a description of the research design, and (f) a description of the data analysis methods.

Rationale for Qualitative Design

While quantitative research design is often referred to as the gold standard of empirical research (Crotty, 1998), there has been always been a place for qualitative research in education. Robert Ebel (1967), who served as president of the American Educational Research Association, wrote:

"The process of education is not a natural phenomenon of the kind that has sometimes rewarded scientific investigation. It is not one of the givens in our universe. It is manmade, designed to serve our needs. It is not governed by any natural laws. It is not in need of research to find out how it works. It is in need of creative invention to make it work better" (p. 81).

Other researchers agree with this assessment. Ross, Morrison and Lowther (2010) believe that research studies on new educational technologies often focus narrowly on proving effectiveness, while failing to address other important issues such as "in what ways, in which contexts, for whom, and why" (p. 31) the technology facilitates learning. They believe that quality educational

research must not only present empirical results on how well a technology application works but also address the issue of why it works (Luo, 2011).

According to Luo (2011), "qualitative perspectives and data collection methods have a long tradition in educational technology research and are gaining more and more attention from researchers in the field" (p. 6). Savnye and Robinson (2004) conducted a search of educational databases and found that in recent years the number of qualitative studies in the field of educational technology has increased dramatically.

Merriam (2009) identifies four characteristics that are critical to qualitative research: (a) the focus is on meaning and understanding, (b) the researcher is the primary source for data collection and analysis, (c) the research process is inductive, and (d) the product is richly descriptive. Merriam states that "all qualitative data analysis is primarily inductive and comparative" (p. 175). The current research study includes all four of these characteristics, and ultimately leads to a comparison of the model under consideration with other evaluation models.

Rationale for Single Case Study Method

Stufflebeam (2001a) believes that case study is an appropriate method to employ in program evaluation. The case study researcher can address issues of accuracy by employing multiple perspectives, information sources, and research methods. Case studies pay particular attention to context and describe contextual influences in depth. A good case study includes systematic analysis of qualitative information. It is designed to focus on the customer's most critical questions.

Merriam (2009) describes single case studies as particularistic. This means that they "focus on a particular situation, event, program, or phenomenon" (p. 43). Stake (1981) claims

that the knowledge gained from a case study is different from other types of research knowledge in four ways. He believes that case study knowledge is:

- More concrete than other types of research knowledge. It resonates with us and our own experience because it is vivid and concrete

- More contextual than other types of research knowledge. Knowledge in case studies is rooted in context, just as our personal experiences are. This is different from the more abstract, formal knowledge of other types of research

- More developed by reader interpretation than other types of research knowledge. The experience and understanding that readers bring to the case study lead to generalizations when new data are added to the case

- Based largely on the reference population determined by the reader. When making the generalizations described above, the reader has a certain population in mind. Therefore, unlike in traditional research, readers participate in extending generalizations to reference populations (Stake, 1981, pp. 35-36)

Setting

Phase One of this study was conducted at a private all-female secondary school in suburban Detroit. The school was founded in 1945. It is operated by a religious order of Roman Catholic nuns. It is part of a system of 69 secondary and elementary schools in the United States operated by the religious order. The school has been accredited by the North Central Association of College and Secondary Schools since 1951. The school is recognized nationally as an Exemplary Private School by the United States Department of Education and the Council for American Private Education.

At the time of the study, the school had 742 students who represent 60 metropolitan communities in the area. The school draws students from a radius of more than 60 miles. Annual tuition at the school is approximately $10,000. 99% of students continue on to college or university. 88% of students attend their first choice of college. The composite ACT score for students in the school is 24.1. The average scores on the SAT are: Critical Reading – 589, Math – 566, Writing – 628.

The staff consists of 89 members. 75% of staff members hold master's degrees or higher. Four Roman Catholic nuns are part of the staff. The average staff member has 16 years of teaching experience. 19 staff members are alumnae of the school, including the principal and the president.

The school initiated a 1:1 technology program in the fall of 2006. During the 2006-2007 school years, all of the freshmen were required to purchase a Hewlett-Packard Elitebook tablet computer. The sophomores, juniors, and seniors were not required, or even allowed, to purchase laptops for use at school. The following school year (2007-2008) the freshmen and sophomores were required to purchase laptops. In 2008-2009, freshmen, sophomores, and juniors were required to purchase laptops. Finally, during the 2009-2010 school years, all four grades were required to have laptops. The students were required to purchase their own laptops at a price of more than $2,000.

The school administration initiated the technology program in response to a mandate from the religious order that owns and operates the school to move the school from "viable to vital". At the time, the school was the only private secondary school in the state to initiate a technology program.

Participants

The participants in the program evaluation included representative stakeholders from the secondary school community. A representative stakeholder is defined as anyone who has the ability to affect or be affected by a program. The stakeholders who participated in the evaluation included: the school president, the principal, the assistant principal, teachers, board members, a representative of the religious order that owns and operates the school, current students, alumnae, current school parents, prospective school parents and the staff members that support the laptop hardware within the school. More than thirty people participated in the evaluation and subsequent metaevaluation of the technology program.

The metaevaluation of the IEP was conducted by several groups. The group of more than 30 people who participated in the evaluation of the technology program also participated in the metaevaluation. In addition, I hired Dr. Carl Hanssen to conduct a metaevaluation of the evaluation. Dr. Hanssen has a PhD in Evaluation, Measurement, and Research and works as a professional evaluator. He has particular expertise in metaevaluation (Hanssen, Lawrenz, and Dunet, 2008).

Research Design

The research design included 2 stages. During the first stage, I used the IEP to conduct an evaluation of a 1:1 technology program. The second stage consisted of a thorough metaevaluation of the evaluation and the evaluation model. Consensus has not been reached within the field of evaluation about the best way to empirically evaluate an evaluation model. Fitzpatrick et al. (2011) believe that there are three groups that can conduct metaevaluations, including the evaluator, the evaluation participants, and a professional evaluator. Fitzpatrick et al. (2011) propose that it is possible to evaluate the design of an evaluation by obtaining an operationalized version of the design, determining who will conduct the metaevaluation,

obtaining permission to conduct the metaevaluation, selecting the appropriate standards by which the evaluation design will be judged, and using those standards to make a final judgment about the adequacy of the design. Miller (2010) proposes a slightly different framework for empirically exploring how well a theory translates into practice. Miller recommends developing or obtaining an operationalized version of the theory, determining the range of application of the theory as well as its feasibility in practice, examining the discernible impact that the theory has in practice, and determining whether or not the impact of the theory is reproducible over time, occasions, and evaluators. Stufflebeam and Coryn (2012) advocate the application of the Joint Committee on Standards for Educational Evaluation's *Program Evaluation Standards (*Yarbrough et al., 2011) to the model under consideration, based on the results of actual evaluations conducted using the proposed model. In the absence of consensus on the best way to evaluate the Impact Evaluation Process, I opted for an "all of the above" approach. This research project addresses each of the methods listed above.

In order to gain the insight necessary to adequately operationalize the model and to use an actual evaluation as the basis for the application of the *Program Evaluation Standards* (Yarbrough et al., 2011), I conducted a complete evaluation using the IEP. Throughout the evaluation process I worked iteratively to develop and fine-tune an operationalized version of the IEP. I worked with Dr. Guerra-Lopez to ensure that I captured the essence of the model in the operationalized version.

Technology program evaluation

Phase One of the study consisted of using the Guerra-Lopez IEP to evaluate the 1:1 technology program at a secondary school. I initiated the process by identifying a local secondary school that implemented a 1:1 technology program in 2006. I approached the

president of the school and asked if she would be interested in working with me to evaluate the program. She consulted with her administrative team and agreed to participate and to give me full access to all potential stakeholders in the program. A letter of support appears in Appendix C. The timing of the evaluation was fortuitous because the school had just announced that they were planning to switch their technology platform from HP tablet computers to iPads. The administration felt that an evaluation of the results that they had achieved in the first six years of the program would inform the transition to iPads. The resulting evaluation was conducted from April to September, 2012. In addition, all of the evaluation participants agreed to participate in a metaevaluation of the evaluation.

The IEP consists of the following steps: (1) identify stakeholders and expectations, (2) determine key decisions and objectives, (3) identify measurable indicators, (4) identify data sources, (5) select data collection methods, (6) select data analysis tools, and (7) communication of results and recommendations. As soon as I received approval from the Wayne State University Human Investigation Committee (Appendix D), I began to work through the steps in the model.

Step 1. Identify stakeholders and expectations. I met with the evaluation client and began to learn about the organization and about the program that I was going to be evaluating. I worked collaboratively with the president, principal, and vice principal of the school to identify all stakeholder groups and to clarify their expectations for the evaluation. I invited representatives of each stakeholder group to participate in the evaluation and clarified the expectations for the evaluation. I developed an informal agreement with the school about the goals, processes, schedule, and reporting procedures for the evaluation.

Step 2. Determine key decisions and objectives. I explained the IEP to all of the representative stakeholders and asked them to help identify the desired results that they sought from the 1:1 technology program at each of the three levels identified by Guerra-Lopez (2007b): societal, tactical, and operational. Based on these desired results, I developed the evaluation questions and reached consensus with the participants about them.

Step 3: Derive measurable indicators. I developed a tentative list of measurable indicators that pointed towards the desired results at each level, and reached consensus with the participants about them. Each of the indicators of the desired results was aligned with the specific evaluation questions and was relevant, reliable, and valid. Based on feedback from the representative stakeholders, I finalized the list of indicators.

Step 4: Identify data sources. With the list of measurable indicators that would point towards the desired results of the 1:1 technology program, I began to identify potential data sources. I worked primarily with the school administration to identify these sources. I made every effort to use existing information where possible in order to avoid duplication in my data collection processes.

Step 5: Select data collection instruments. Based on the measurable indicators and the identified data sources I developed a list of data collection instruments that would ensure that the data was collected efficiently and accurately. I developed and deployed several surveys in order to collect information about certain stakeholder groups' perceptions of the technology program. The school provided me with a variety of data, such as ACT scores, college scholarship information, college choice information, endowment/donation information, and enrollment data. The data was collected throughout the summer of 2012.

Step 6: Select data analysis tools. I analyzed the data in a variety of ways, depending on the type of information and the format in which it existed. Based on the analysis of the data, I addressed each of the evaluation questions, and arrived at conclusions about whether the organization had achieved the results that they sought when the program was created. I developed a set of recommendations for the future based on my findings.

Step 7: Communicate results and recommendations. I developed a written report based on the analyzed data and presented it to the school administration. I created a video presentation that I emailed to the representative stakeholders with a copy of the preliminary report. Based on feedback that I received from the stakeholders, I made small revisions to the report and issued a final version of it in September, 2012. I offered to facilitate the implementation of the report's recommendations.

Metaevaluation

Stufflebeam & Coryn (2012) conducted a metaevaluation of nine evaluation models in an attempt to help evaluators appraise their relative merits. They based their metaevaluation on their personal experience with applications of each approach (Stufflebeam & Coryn, 2012). They each used Stufflebeam's (2011) *Checklist* to arrive independently at a score for each evaluation approach, and then they compared scores and reached consensus on a final score for each model.

The *Checklist* contains ratings for each of the thirty evaluation standards. Each standard is broken down into six checkpoints, which are represented by individual statements. In total, there are 180 statements that make up the metaevaluation *Checklist*. The statements refer to desired elements of each evaluation standard. If the element is present in the evaluation, the person filling out the checklist places a checkmark next to it. This is a useful technique which simplifies the judgment of each standard by reducing it to a dichotomous scale.

Next, the evaluator counts the number of checkmarks within each standard. If the evaluator checked all six statements within the standard, it is assigned a rating of "Excellent". If the evaluator checked five of the six statements, the standard is assigned a rating of "Very Good". If four of the six statements are checked, the standard is assigned a rating of "Good". If two or three of the six statements are checked, the standard is assigned a rating of "Fair". If none of the statements are checked, the standard is assigned a rating of "Poor".

The rating for each standard is then weighted, with the number of "Excellent" ratings multiplied by 4, the number of "Very Good" ratings multiplied by "3, the number of "Good" ratings multiplied by 2, and the number of "Fair" ratings multiplied by 1. The weighted scores are then summed to obtain a total score for each Category. The total score is divided by the maximum number of points available for that category and multiplied by 100 to arrive at a percentage score. These scores vary by category and are shown below in Table 5.

Table 5 Stufflebeam (2011) Metaevaluation Checklist Scores by Category

Category	# of Standards in Category	Scoring	Rating
Utility	8	29.44-32 (92-100%)	Excellent
		21.44-29.43 (67-91%)	Very Good
		13.44-21.43 (42-66%)	Good
		5.44-13.43 (17-41%)	Fair
		0-5.43 (0-16%)	Poor
Feasibility	4	14.72-16 (92-100%)	Excellent
		10.72-14.71 (67-91%)	Very Good
		6.72-10.71 (42-66%)	Good
		2.72-6.71 (17-41%)	Fair
		0-2.71 (0-16%)	Poor
Propriety	7	26.76-28 (92-100%)	Excellent
		18.76-26.75 (67-91%)	Very Good
		11.76-18.75 (42-66%)	Good
		4.76-11.75 (17-41%)	Fair
		0-4.75 (0-16%)	Poor

Accuracy	8	29.44-32 (92-100%)	Excellent
		21.44-29.43 (67-91%)	Very Good
		13.44-21.43 (42-66%)	Good
		5.44-13.43 (17-41%)	Fair
		0-5.43 (0-16%)	Poor
Evaluation Accountability	3	11.04-12 (92-100%)	Excellent
		8.04-11.03 (67-91%)	Very Good
		5.04-8.03 (42-66%)	Good
		2.04-5.03 (17-41%)	Fair
		0-2.03 (0-16%)	Poor

Finally, a score for overall merit is calculated in two ways. First, a score is calculated assuming that each category is equally important. This is calculated by averaging the scores for each of the five categories. Alternatively, each category is weighted for importance (Utility=.23, Feasibility=.09, Propriety=.25, Accuracy=.34, and Evaluator Accountability=.09) and a weighted score for overall merit is calculated.

Data analysis

Research Questions 2-6 were addressed by applying Stufflebeam's (2011) *Checklist* from several perspectives. At the beginning of this research project, the representative stakeholders who participated in the technology program evaluation also agreed to participate in the metaevaluation. After I completed the first three steps of the IEP, as described in the previous section, I asked the stakeholders to take an online survey as the first part of the metaevaluation. The survey consisted of 75 statements that were taken directly from Stufflebeam's (2011) *Checklist*. Each of the questions addressed the evaluation *process*. After the final technology program evaluation report was completed and delivered to the representative stakeholders, I asked them to take a second online survey as the second part of the metaevaluation. The second

survey consisted of 53 questions that were also taken directly from the Stufflebeam's (2011) *Checklist*. Each of the questions in the second survey addressed the evaluation *findings and report*. Taken together, the two surveys asked the stakeholders to address 128 of the 180 items in Stufflebeam's (2011) *Checklist*. The remaining 52 items were deemed inappropriate for the stakeholders to address because the items required a certain level of expertise in evaluation that I could not assume the stakeholder participants possessed. For example, I did not ask the stakeholders to answer questions about the evaluation's validity, reliability, explicit reasoning, or internal metaevaluation. Of the 30 representative stakeholders who participated in the program evaluation, 20 completed both online surveys. I used the mode of stakeholder responses to represent the score for the stakeholder group for each standard. The mode represents the most common response among the stakeholders and is appropriate because of the nominal nature of the data. I plugged the stakeholder responses into Stufflebeam's scoring formula to calculate an evaluation score for the stakeholder group.

After the evaluation was completed and delivered to the client, I filled out Stufflebeam's (2011) *Checklist* based on my personal assessment of the evaluation. I also hired Dr. Carl Hanssen, an independent evaluation consultant, to review the evaluation and conduct a metaevaluation. Dr. Hanssen is a graduate of the Western Michigan University Evaluation program and is familiar with the Stufflebeam's (2011) *Checklist*. Dr. Hanssen's agreement to conduct the metaevaluation is attached in Appendix E. I sent a copy of the evaluation report, along with additional documentation of my work, to Dr. Hanssen. He reviewed the material, completed the *Checklist,* and calculated a score for the evaluation. After reviewing his preliminary report, I offered further documentation on certain aspects of the evaluation and Dr. Hanssen revised his score.

In order for metaevaluation to be effective, it has to provide valid information. Validity in research is the degree to which the data gathering processes and procedures measure what they are intended to measure (Furr & Bacharach, 2008; Joint Committee, 2010). In the context of metaevaluation, validity refers to the extent "to which the process of gathering information about an evaluation and applying criteria to that information yields accurate conclusions about that evaluation's quality" (Wingate, 2009, p. 39). A critical prerequisite for validity is reliability. Reliability refers to the consistency of a tool across contexts (Juni, 2007). While there are several different types of reliability, within the context of metaevaluation interrater reliability is of primary interest. Interrater reliability refers to the consistency of a tool when applied to the same object by different individuals (Juni, 2007).

There are several ways to calculate interrater reliability. Correlational techniques measure consistency, but are not adequate to measure actual agreement. Percent of agreement is the simplest measurement technique and simply consists of counting the number of total agreements between raters and dividing that by the number of possible agreements (Wingate, 2009). However, this approach does not account for agreement due to chance. Cohen's Kappa statistic is a measure of agreement that factors chance into the equation. Unfortunately, it is not appropriate for this study because it is limited to calculating reliability for two raters. In this study there were 22 raters. Fleiss' kappa (1971) accounts for any constant number of raters giving categorical ratings to a fixed number of items. The formula returns a value between zero and one. Landis & Koch (1977) interpret the significance of Fleiss' kappa as seen in Table 6 below. I am particularly concerned with interrater reliability because this study is based on shared judgment of criteria, and a higher level of agreement strengthens the basis for accepting the shared judgment as reliable and valid. Stufflebeam & Coryn (2012) based their evaluation of the models

on their own expertise, and since the evaluation participants and I do not share this same level of

expertise, high interrater reliability helps to triangulate and strengthen the results.

I calculated the value of Fleiss' kappa for this study in two ways. First, I averaged the

ratings of the 20 participants in the program evaluation who completed both parts of the online

survey to arrive at a single value for the stakeholder group. I compared this result with my results

and the results from the professional metaevaluator. Nineteen of the 30 standards on

Stufflebeam's (2011) *Checklist* were rated by all three groups. Second, I calculated Fleiss'

Kappa by counting each stakeholder's responses individually, which increased the number of

raters from three to 22. Landis & Koch's (1977) interpretation of the significance of κ appears in

Table 6.

Table 6 Interpretation of Fleiss' Kappa

κ	Interpretation
<0	No agreement
0.0-0.19	Poor agreement
0.20-0.39	Fair agreement
0.40-0.59	Moderate agreement
0.60-0.79	Substantial agreement
0.80-1.00	Almost perfect agreement

Summary

In this chapter, I described the methodology that I used to empirically examine the

Impact Evalution Process (IEP). I conducted an evaluation of a 1:1 technology program at a

secondary school using the IEP. During this evaluation, I operationalized the IEP in order to

begin to build evidence of a connection between theory and practice. At the conclusion of the evaluation, I used Stufflebeam's (2011) *Checklist* to conduct a metaevaluation. I also asked the participant stakeholders to complete the majority of the *Checklist.* Finally, I hired a professional metaevaluator to conduct a metaevaluation using the *Checklist.* Based on the three groups' metaevaluation scores, and taking into consideration their level of interrater agreement, I began to develop a summarizing judgment about the effectiveness of the IEP for use in evaluating educational programs, and specifically judged its effectiveness against the Joint Committee's *Program Evaluation Standards (*Yarbrough et al., 2011). The following chapter discusses the results of the study.

CHAPTER 4 Results

Introduction

This chapter discusses the results of the study described in Chapter 3. The goal of this study was to empirically evaluate the Guerra-Lopez Impact Evaluation Process (IEP). Since there is no consensus in the field of evaluation about the best way to empirically evaluate an evaluation model, this study used several processes recommended by leaders in the field to evaluate the model in order to triangulate and strengthen the results.

The chapter is divided into four sections. In the first section, I describe placement of the IEP in the body of evaluation research. The second section describes the results I obtained by following the empirical evaluation frameworks recommended by Miller (2010) and Fitzpatrick et al. (2011). In the third section, I describe the results of the program evaluation that I conducted as part of the metaevaluation process. Finally, I describe the application of the Stufflebeam (2011) *Checklist* to the program evaluation, and the resulting analysis of the research questions. Each of the sections informs the subsequent sections. This chapter ultimately addresses the following research questions:

Research Question 1: Where does the Guerra-Lopez IEP fit into the body of evaluation research?

Research Question 2: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Utility?

Research Question 3: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Feasibility?

Research Question 4: To what extent does the IEP meet the Joint Committee Program Evaluation requirements for Propriety?

Research Question 5: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accuracy?

Research Question 6: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accountability?

Research Question 7: How effective is the IEP for evaluating educational programs?

Placement of the IEP in Body of Evaluation Research

Scholars within the field of evaluation use a wide variety of approaches for determining the merit or worth of programs under consideration. Objectivists are likely to be drawn to quantifiable methods and results while constructivists often seek to understand effectiveness as it is perceived by users. Scholars with a utilitarian orientation will concentrate on measuring the overall impact of a program on all stakeholders, while scholars that ascribe to an intuitionist-pluralist approach are more likely to be concerned with a program's impact on each individual. This variety of philosophical and methodological differences has led to a wide range of frameworks for classifying evaluation theories. In an attempt to identify the correct placement of the IEP within the field of evaluation, I examined three of these frameworks as part of this research study: Fitzpatrick et al. (2011), Stufflebeam & Coryn (2012), and Christie & Alkin (2012).

Fitzpatrick et al. (2011) divide all evaluation approaches into four categories: program oriented, decision oriented, consumer and expertise oriented, and participant oriented. The IEP is most similar to the decision oriented approach, which holds that the purpose of evaluations should be to provide decision makers with information that will help them make sound decisions.

The decision-oriented approach is closely related to the systems approach and provides clear focus to evaluations as well as stresses the importance of the utility of information.

Stufflebeam & Coryn (2012) classify all evaluation approaches in four categories: questions and methods, improvement and accountability, social agenda and advocacy, and eclectic. Based on this schemata, the IEP falls into the eclectic category alongside Patton's Utilization-focused model. The two models share many characteristics, including a primary concern with impact and the input of representative stakeholders to focus the evaluation. They both rely on mixed methods and take great pains to ensure that the results are part of a continuous improvement process.

Christie & Alkin (2012) use a framework that is depicted graphically as a tree, and includes the following categories: methods, values, and use. Within this framework the IEP is appropriately placed on the Use branch, near Stufflebeam's CIPP model and Patton's Utilization approach. The Use branch is largely based on Stufflebeam's work, and is strongly oriented towards decision-making. The concept of "use" in evaluation has expanded to include a consideration of the evaluation's influence and capacity to affect change within an organization.

The analysis of the placement of the IEP within the body of evaluation addresses Research Question 1 "Where does the IEP fit into the body of evaluation research?" It also highlights critical features of the model. These include a fundamental concern with the impact that the evaluation will have on organizational results, clients, and society as well as the key role that representative stakeholders play in focusing the evaluation.

Operationalizing the IEP

Fitzpatrick et al. (2011) suggest the following steps for conducting a metaevaluation of an evaluation design: (1) Prepare a copy of the design in a form that is ready for review, (2) Clearly identify who will do the evaluation, (3) Verify that approval has been given to evaluate the design, (4) Apply the appropriate standards to the evaluation design, and (5) Judge the adequacy of the evaluation design. Miller (2010) proposes a similar framework to empirically evaluate the way that evaluation theory informs practice. Miller advocates "operational specificity" as the key to making an evaluation theory useful in practice, and to making its "theoretical signature" recognizable (p. 391). Both of these approaches require an operationalization of the model under consideration.

The IEP consists of seven simple steps: (1) identify stakeholders and expectations, (2) determine key decisions and objectives, (3) derive measurable indicators, (4) identify data sources, (5) select data collection instruments, (6) select data analysis tools, and (7) communicate results and recommendations. In order to operationalize the model, I referred to Guerra-Lopez's (2007a, 2007b, 2007c, 2007d, 2008, Guerra-Lopez & Toker, 2012) writings and tentatively identified sub-steps that explained the activities that are required throughout the process. In order to verify that the IEP conformed to the *Program Evaluation Standards* identified by the Joint Committee on Standards for Educational Evaluation (Yarbrough et al., 2011) and to increase the internal validity of the operationalized model, I identified a link between each step in the IEP and Stufflebeam's (2011) *Metaevaluation Checklist*. As I conducted the evaluation of the 1:1 technology program that is part of this research study, I iteratively reviewed and modified the operationalized steps. I sent a preliminary version of the operationalized IEP to Dr. Guerra-Lopez and she offered suggestions and clarifications. The final operationalized version of the IEP is seen below in Table 7. The operationalized version of the IEP is critical to this study

because it provides a concrete link between theory and practice. If an evaluation is based conceptually on the IEP, and is conducted using an operationalized version that clearly links the process to the *Program Evaluation Standards* (Joint Committee, 2010) then it is possible to begin to make judgments about the IEP's effectiveness as a prescriptive model.

Table 7 Operationalized Version of the Impact Evaluation Process

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
Prior to Beginning:			
	0.1 Hire a qualified evaluator	Hire experienced, qualified, competent evaluation team	U1.1, U1.2, U1.3, U1.4, U1.5, F1.3, A3.3, F2.6, F3.2, A6.5
		Begin to plan metaevaluation	E2.1
1. Identify Stakeholders and Expectations			
	1.1 Identify stakeholders	Learn about the organization, including, political, cultural, economic contexts, history, and institutional protocols	F3.1, P1.1, P3.2,
		Collaboratively identify all stakeholder groups	U2.2, P1.2,
		Invite representatives of each stakeholder group to participate in evaluation	U2.3, P1.3, P1.5
		Engage representative stakeholders in deriving a full description of program to be evaluated based on needs	P5.4, A2.4
	1.2 Identify stakeholder expectations	Identify expectations of each group, including: - What evaluation questions will be answered? - Tie expectations to organizational mission - Clarify values at basis of evaluation - Clarify rules to allocate resources fairly and transparently - Clarify what will be the final products of the evaluation and what criteria they must meet	U4.2, U4.3, U4.4, U4.5, U4.6, P4.1, E3.2, U5.1,
		Neutralize possibility that evaluation will be used as tool of manipulation for special interest groups	P1.6
		Keep focus on results and consequences: -Help stakeholders articulate the results and consequences they seek -Use these results and consequences as basis for negotiated agreements -Use negotiated agreements to monitor progress	P2.5; P2.6

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
		-Ensure adequate links to methods, means, and other "how to's" to expected results and consequences -Update/Revise as required	
		Clarify what is expected of evaluator: - Under what criteria will performance be judged - What will be the communication process - Clarify evaluator's ethical principles, values and code of professional conduct	P3.4, P2.1, P3.3, E2.3, U4.1
		Clarify what will be expected of stakeholders: - What type of support will they be expected to provide? - What type of feedback and how often? - Data collection assistance? - Administrative assistance? - Other assistance? Clarify that stakeholders may participate, withdraw, and challenge decisions throughout evaluation process	P2.1, P3.5,
	1.3 Ensure Commitment from Client	Recognize that client might have mixed feelings about evaluation: - Emphasize commitment to continual improvement - Common purpose and shared destiny are key to getting and maintaining commitment - Listen carefully to stakeholders - Involve stakeholders in setting objectives - Look for ways to build trust	P1.2., P1.3, P1.4, P1.5
	1.4 Determine whether project is viable	Evaluate likelihood that evaluation will not be successful due to unreasonable expectations or political issues. If likelihood is high, decline to pursue evaluation	F2.1
	1.5 Develop contract	Include formal management plan in contract. Include: goals, procedures, communication, schedule, budget, monitoring, risk management, accounting, and reporting	U1.6, U2.1, U7.6, F1.2, F4.1, P2.2, P2.3, P2.4, P2.5, P3.1, A8.1, A8.2, E1.1, E1.2, E1.3, P7.1, U6.1, P7.2
2. Determine Key Decisions and Objectives of Evaluation			
	2.1 Determine what decisions	Stakeholders identify the decisions that they would like to be made with evaluation findings	U2.5

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
	will be made with the evaluation findings.	and recommendations	
	2.2 Determine objectives of evaluation	Stakeholders identify objectives of evaluation	U3.1
		Clarify and justify differences in valuing of certain stakeholders' evaluation needs over those of others	P4.4
		Gain consensus among stakeholders regarding evaluation boundaries	U2.4
	2.3 Explain Organizational Elements Model to stakeholders	- Recognize that all organizations have ultimate results that they want to achieve, as well as building block results that have to be accomplished en route - Identify ultimate goal for organization – ideal outcomes/impact on community and society - Identify organization-level results/outputs through which organization ultimately seeks to reach Ideal Vision - Identify internal building-block results that, when properly linked, deliver the mission	U3.2, A6.1,
	2.4 Identify results sought from program being evaluated at societal, tactical, and operational levels	Facilitate stakeholder identification of the results that are sought at each level	U3.3, U3.4, U3.6
		Gain consensus on identified results	U6.2
	2.5 Articulate evaluation questions that are tied to desired results	First question to ask: How much closer to our Ideal Vision and our mission did we get as a result of this program?	
		Identify other questions at tactical and operational levels. Questions should be aligned with results. Are we achieving the results we want to achieve at each of these levels?	
		Finalize consensus on evaluation purpose, questions, procedures, and their underlying rationale	P4.3
3. Derive Measurable Indicators			
	3.1 Evaluator develops list of measurable	Identification of measurable indicators should be comprehensive enough to assure adequate scope for assessing program merit, worth, and	U5.3

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
	indicators (observable phenomena) for each desired result	significance	
		Indicators of desired results should be reliable, relevant, valid, and complete	A1.1
	3.2 Stakeholder Review, Feedback, and Revision	Provide stakeholders with new information as it emerges and get appropriate feedback. Revise as required	U2.6, U6.3, U7.2
		Stay within boundaries of stated purpose and evaluation questions	U5.4
4. Identify Data Sources			
	4.1 Evaluator identifies potential data sources	Selected data sources should provide reliable, relevant, valid, and complete information on identified measurable indicators	U5.2, A5.1, A6.2, A6.3
		Use existing information where possible to avoid duplication in collecting data	F2.5
	4.2 Stakeholder Review, Feedback, and Revision	Provide stakeholders with new information as it emerges and get appropriate feedback Revise as required	F2.3, U5.6, F2.4, U2.6, U6.3, U7.2
5. Select Data Collection Instruments			
	5.1 Evaluator identifies methods and instruments to collect data from identified data sources	Choose data collection methods that will cause minimal disruption while providing sufficient scope and depth of information	F2.2, F3.4, U5.5, E2.2
	5.2 Evaluator Collects Data	Ensure that data collection is systematic, replicable, and well documented	A2.2, A2.3
		Retain and store data securely	A5.2, A5.3, A5.4, A5.5, A5.6
	5.3 Stakeholder Review, Feedback, and Revision	Provide stakeholders with new information as it emerges and get appropriate feedback Revise as required	
6. Select Data Analysis Tools			
	6.1 Evaluator	Ensure validity and reliability of data	A2.5, A3.1, A3.6,

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
	determines best methods to analyze all varieties of data	Develop plan for analyzing data	A6.4, A6.6
	6.2 Evaluator Analyzes Data	Use generally accepted accounting and auditing practices	P7.3, P7.6
		Maintain accurate and clear records	P7.4, P7.5
	6.3 Evaluator Derives Conclusions	Base conclusions on relevant, reliable, valid, and complete data	A1.2
	6.4 Stakeholder Review, Feedback, and Revision	Continue to communicate progress and findings as they emerge and get appropriate feedback Revise as Required	XP4.5
7. Communicate Results and Recommendations			
	7.1 Emphasis is on confirming that results obtained are ones initially intended when program was implemented	Avoid premature judgments	
		Develop appropriate and feasible recommendations in conjunction with stakeholders	
		Provide stakeholder with preliminary report and ask for feedback. Revise as required	U6.5, P5.3
		Discuss confidentiality and privacy policies	A8.5, A8.6, E1.4
	7.2 Confirm type and format of report required based on initial expectations and contractual obligations	Alternatives include report based on: evaluation results, alternative courses of action, goals & objectives, performance records	
		Include evaluation's costs in report	F4.5, P5.6
		Prepare and deliver report in timely manner	U7.1
		Report should include description of all important aspects of the program so as to be independently replicable	XA4.1, A4.2, A4.4, A4.5, A4.6
	7.3 Know the audience	Use clear and concise language that stakeholders will understand - descriptive, non-judgmental	U7.4, A1.4
		Format report based on stakeholder expectations and preferences, employing multiple reporting mechanisms, as needed	U7.3, U8.1, A8.3, A8.4
		Identify conflicts of interest and how they were resolved	P6.6, P6.1, P6.2, P6.3, P6.4, P6.5
	7.4 Identify the Key Message:	Include discussion of who should take action and what the action should be in recommendations	

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
	Take Action!	Guiding question should be: What did the data consistently point to and what is the most likely way to effectively and efficiently deal with it? What are the costs and consequences of these recommendations?	U8.4, F4.6, A2.6, A7.2
	7.5 Make clear distinctions and linkages among recommendations interpretations, findings and analyzed data	Do not confuse interpretation of data with what the data itself indicates	U8.5
		Limit conclusions to applicable purpose, context, and activities based on initial stakeholder expectations and contractual agreements	A1.3
		Identify important assumption, procedures, interpretive frameworks, models, values, and caveats	A1.5, A3.4, A3.5, A4.3, A7.1, A7.3, A7.4
		Clearly articulate evaluation's conclusions and limitations	P5.5, A7.6
		Identify alternative explanations and reasons for rejecting them	A1.6, A3.2, A7.5
	7.6 Be clear about responsibility of stakeholders	Clarify link between results and each level (strategic, tactical, and operational)	
		Encourage "internal locus of control" mentality to empower stakeholders to implement feasible recommendations	
		Consider additional deliverable of a general project management plan for the implementation of recommendations, as appropriate given context, and contractual obligations	
		Articulate constraints under which evaluation information can be released	P5.1
	7.7 Clarify Evaluator's Role After the Report	Plan follow-up activities, if needed, to assist stakeholder with interpretation and use of final report recommendations	U7.5, U8.3, U8.6, F3.5,
	7.8 Conduct metaevaluation of report	Explain need for external assessment of evaluation	E3.1
		Retain relevant documents for use in metaevaluation	E1.5, E1.6, E2.4, E3.5
		Retain qualified metaevaluator	E2.3, E3.3, E3.4, E3.6
		Make results of metaevaluation available to stakeholders	E2.5, E2.6
Throughout the Process			
	Iteratively:	Continuously manage evaluation based on	F1.1

Step #:	Operationalized Steps in Process:	Sub-steps in IEP	Related Stufflebeam *Checklist* Steps
		stakeholder needs and evaluation's purpose	
		Communicate with stakeholders throughout process	F1.4, U6.4, A2.1, U8.2, F3.6, P1.4, P5.2
		Systematically oversee and document evaluation's activities and expenditures	F1.5, F4.2, F4.3,
		Periodically review and update evaluation's purpose, process, procedures, and reports to assure alignment with stakeholder needs	U3.5, U3.6, U6.6, F1.6, F3.3, P2.6, P3.6,
		Assure that purposes, questions, procedures, and findings are transparent and accessible to all stakeholders	P4.2

1:1 Program Evaluation

In order to prepare for the metaevaluation of the IEP, and as part of the process of operationalizing the model, I used the IEP to conduct an evaluation of a 1:1 technology program in a secondary school. This constituted Phase One of the study. In this section, I describe the results of the technology program evaluation.

Step 1: Identify stakeholders and expectations. As part of the first step in the evaluation, I met with the president, principal, and assistant principal of the school to preliminarily identify the stakeholder groups who would participate in the evaluation. We agreed that the stakeholder groups included the administration, faculty, board of directors, members of the religious order that operate the school, current students, recent graduates (who were the first to participate in the 1:1 technology program), current school parents, and the school staff that supports the technology program.

Based on this first meeting I began to identify individuals from each of these stakeholder groups who were interested in participating in the evaluation process. The majority of the group

is best described as a convenience sample since they were not chosen entirely at random. The school administration suggested people from each group. I found several participants through informal networking at school functions. I invited several members of the board of directors, based on the administration's recommendations. I asked faculty members to recommend current students and recent graduates who would be willing and able to participate and who would provide thoughtful insight. I invited the two members of the religious order who are most active in the school. I invited all of the technology support staff to participate.

During my first meeting with the administration they explained to me that the 1:1 technology program was not uniformly popular with the faculty at the school. Some members of the faculty have embraced the use of technology in their classrooms while others have resisted it emphatically. In order to get fair representation of the faculty in the evaluation, I decided to randomly select individuals to participate. However, since I wanted to ensure that each academic department was represented, I stratified the sample by department. I randomly chose three members of each academic department and invited them to participate. Ultimately, one member of each department agreed to participate.

I spoke to most of the participants on the phone or in person to explain the project, and followed up the conversation with an email that explained that their participation would include providing their input as part of an evaluation of the technology program as well as providing feedback as part of a metaevaluation. A copy of the email can be found in Appendix F. All participants were required to sign a participation agreement. The current students under the age of 18 were required to obtain written parental permission to participate in the research project. A copy of the informed consent form can be found in Appendix G. The email explained the research project as well as what would be required of participants. As a token of appreciation for

their participation, I offered current students and recent graduates an iTunes gift card. Adult participants were entered into a drawing for an eReader.

After the initial groups of stakeholders were identified, I asked them to help me determine if there were any other potential stakeholders in the evaluation program. One of the participants suggested that I should invite people who were in the process of choosing a high school for their daughter to join in the evaluation because of the possibility that the existence of a 1:1 technology program might influence their decision. I identified prospective families from a list of Open House attendees and one of the parents joined the evaluation.

Once the participants were identified and registered, it was critical that they learn about the Guerra-Lopez Impact Evaluation Process. The participants were geographically diverse, so I created a PowerPoint presentation that introduced them to the IEP and to the evaluation project. Copies of the PowerPoint slides can be found in Appendix H.

Step 2: Determine key decisions and objectives. The second step of the IEP is determining the key decisions and objectives of the evaluation. This process began during my preliminary meeting with the school administration. They indicated that they were committed to continuing the program regardless of the evaluation's findings. They were open, however, to making adjustments to the program if the evaluation indicated that they needed to be made. They were also committed to transitioning to iPads, regardless of the evaluation's findings. I made it clear to all of the participants that these two issues would not be considered as part of the evaluation.

As part of the second step of the evaluation, I asked all of the stakeholders to answer the following question "What did we (the school community) hope to accomplish by starting the laptop program?" The IEP attempts to answer this question by identifying intended results at

three levels: strategic, tactical, and operational. Over the course of several weeks, I met with the evaluation participants to address this question. The administration provided a wealth of information that helped to answer this question. One of the most informative artifacts that they provided is the "Technology Vision Statement" which was developed in 1998, and which served as the key foundational element as the school developed their plan for a 1:1 technology program.

Through meetings and email exchanges with representative stakeholders, I collected extensive feedback about the intended results of the technology program. I sorted the accumulated information into three groups, based on the three levels of results (strategic, tactical or operational) that the model addresses. I coded the sorted feedback by theme and developed a preliminary list of the results that the stakeholders were interested in. I sent this preliminary list to all of the participants and asked for feedback. Based on the responses that I received, I developed a final list of the results that the representative stakeholders were seeking from the technology program and distributed them to the participants. The evaluation questions that emerged through this process are:

1. Has the program helped students to develop college and 21st Century skills?

2. Has the program increased students' competence and confidence in the creation and use of information?

3. Has the program enriched the classroom experience and increased motivation by making learning more enjoyable and engaging students in their own knowledge creation and use?

4. Has the program helped the faculty to implement the curriculum more effectively and facilitated the individualization of instruction?

5. Has the program facilitated communication between students and faculty?

6. Has the program helped WXYZ High School distinguish itself from other schools in the area and has it facilitated marketing?

7. Has the program helped WXYZ High School move from "viable to vital"?

8. Has the program incorporated the global shift in learning towards the concept of "knowledge creation and use by the individual"?

9. Has the program extended the boundaries of the learning environment beyond the school building and 8 a.m. – 3 p.m.?

10. Has the program promoted the integration of values into teaching and learning?

11. Has the program helped WXYZ High School to educate women who make a difference and foster the spiritual, intellectual, moral, physical, and cultural development of its students?

Step 3: Derive measurable indicators. For each evaluation question listed above, I developed a list of potential measurable indicators. Measurable indicators are "observable phenomena that are linked to something that is not directly observed and can provide information that will answer an evaluation question" (Guerra-Lopez, 2007a, p. 16). I distributed the preliminary list of measurable indicators to all of the representative stakeholders and asked them to provide feedback and additional ideas.

Step 4: Identify Data Sources. For each measurable indicator that was identified above, I identified a potential data source. These are listed below in Table 8. Many of the data sources were readily available, including the ACT scores, PSAT scores, enrollment data, college scholarship and attendance data, and professional development information. There were several indicators, however, that could not be readily observed with existing data sources. These included perceptions about the technology program by current students, recent graduates, faculty members, and parents.

Table 8 Measurable Indicators and Data Sources for Each Evaluation Question

Evaluation Question	Measurable Indicators	Data Sources
1. Has the program helped students to develop college and 21st century skills?	• Changes in standardized test scores • Current students' perceptions • Recent graduates' perceptions	• ACT Scores/10 years • Current Students • Recent Graduates
2. Has the program increased students' competence and confidence in the creation and use of information?	• Current students' perceptions • Recent graduates' perceptions • Implementation of challenge-based learning	• Current Students • Recent Graduates • Faculty
3. Has the program enriched the classroom experience and increased motivation by making learning more enjoyable and engaging students in their own knowledge creation and use?	• Current students' perceptions • Recent graduates' perceptions • Faculty perceptions • Implementation of challenge-based learning	• Current Students • Recent Graduates • Faculty
4. Has the program helped the faculty to implement the curriculum more effectively and facilitate the individualization of instruction?	• Changes in standardized test scores • Faculty professional development • Slice impact – lower and higher students should benefit most • Faculty perceptions	• Standardized Test Scores • Professional Development Schedule • Faculty
5. Has the program facilitated communication between students and faculty?	• Current students' perceptions • Faculty perceptions • Use of technology based c0mmunication tools	• Current Students Faculty
6. Has the program helped WXYZ High School distinguish itself from other schools in the area and has it facilitated marketing?	• Computing program as a factor in attracting new students • Computing program at WXYZ High School compared to similar schools	• Parents • Current Students • Extant data re: competitor schools
7. Has the program helped WXYZ High School move	• Enrollment trends • Market share compared to	• Enrollment data for all schools in area

from "viable to vital"?	similar schools • Donation/Endowment trends • Maintain competitive/affordable tuition • Changes in standardized test scores	• Donation data/10 years • Tuition rates for all schools in area • ACT scores/10 years
8. Has the program incorporated the global shift in learning towards the concept of "knowledge creation and use by the individual"?	• Implementation of challenge based learning • Faculty development and use of unique technology-based teaching methods • Level of technology integration in all areas of curriculum	• Faculty • Current Students
9. Has the program extended the boundaries of the learning environment beyond the school building and 8 a.m. – 3 p.m.?	• Faculty use of unique technology-based teaching methods • Level of technology integration in all areas of curriculum	• Current Students • Faculty
10. Has the program promoted the integration of values into teaching and learning?	• Movement towards green technologies such as textbook-less classrooms • Development and application of technology acceptable use policies and cyberbullying policies	• Extant data from school administration
11. Has the program helped WXYZ High School to educate women who make a difference and foster the spiritual, intellectual, moral, physical, and cultural development of its students?	• Number of scholarships awarded to graduates • College attendance rates • Quality of colleges attended • Number of students who pursue majors/careers in Science, Technology, Engineering, or Math (STEM fields)	• Scholarship data/10 years • College attendance data/10 years • PSAT data/10 years

Step 5: Select Data Collection Methods. The primary data collection method that I used

was to obtain existing reports from the school administration, including: an internal survey of

faculty conducted in 2011, ACT scores, PSAT Scores, college scholarship data, college choice data, the technology department professional development schedule, donation/endowment data, the student agenda book. I also studied the websites for competitor schools in the area, and requested secondary school enrollment data from the archdiocese in which the school is located.

In order to collect information from the parents, current students, recent graduates, and the faculty, I created surveys designed specifically to answer the evaluation questions. Copies of these surveys can be found in Appendix I. The paper-and-pencil parent survey was mailed to 350 households during the summer of 2012. I received 144 responses from a population of approximately 700 (21.8% response rate). These results compute to a 7% margin of error at 95% confidence level. The current student, recent graduate, and faculty surveys were conducted online and were sent via email addresses supplied by the school administration. The current student survey was sent to 750 students. With 209 respondents (27.9%), the survey results compute to 5.8% margin of error at 95% confidence. The recent graduate survey was sent 221 students. With 69 responses (31.2%), there is an 8.8% margin of error at 95% confidence level. The faculty survey was sent to 45 classroom teachers. With 34 responses (75.6%), there is an 8.5% margin of error at 95% confidence level. The stakeholders were satisfied with these margins of error for the purposes of this evaluation.

Step 6: Select data analysis tools. Each evaluation question called for its own data analysis tools. For example, question #1 (Has the program helped students to develop college and 21st century skills?) was analyzed using a variety of tools. College readiness skills were analyzed by creating a graph of the past ten years of ACT test scores and looking for a trend in the scores. Written and oral communication skills were identified as college readiness skills by the stakeholders and were analyzed by graphing the past ten years of PSAT writing skills scores.

These skills were further analyzed by summarizing the recent graduates' perceptions of the effect of the laptop program on their written and oral communication skills. Twenty-first century skills were defined by the stakeholders as: the ability to access information, the ability to analyze information, organizational skills, and critical thinking skills. These data were analyzed by summarizing current students' and recent graduates' perceptions of their skills in these areas as well as by graphing PSAT Critical Reading test scores for the past ten years. Each of the other 10 questions in the evaluation was similarly analyzed.

Step 7: Communicate Results and Recommendations. After all of the data were collected and analyzed, I wrote a summary report of my findings. The report consisted of an executive summary, a detailed analysis of each evaluation question, and recommendations for the future. A copy of the final report is included in Appendix J. I met with the administration to present the findings. I recorded a video presentation summarizing the findings and distributed it, along with a hard copy of the report, to all of the evaluation participants. I offered to support the school in the implementation of the evaluation's recommendations.

Application of Stufflebeam & Coryn (2012) *Metaevaluation Checklist*

After I completed Step 3 of the IEP in the technology program evaluation, I asked the representative stakeholders who were participating in the evaluation to complete an online survey. The survey consisted of 75 statements that were taken directly from Stufflebeam's (2011) revised *Program Evaluation Metaevaluation Checklist* (see Table 2). Each of the questions addressed the evaluation *process*. After the final technology program evaluation report was completed and delivered to the representative stakeholders, I asked the representative stakeholders to take a second online survey as the second part of the metaevaluation. The second survey consisted of 53 questions that were also taken directly from the Stufflebeam (2011)

Checklist. In total the representative stakeholders addressed 128 of the 180 checkpoints on the survey. The remaining checkpoints required expertise, or at least familiarity, with evaluations and were potentially beyond the participants' scope of knowledge. Of the 30 people who participated in the evaluation, 20 completed both checklists for a 70% response rate. All stakeholder groups were represented in the metaevaluation with the exception of the religious order that owns the school. Neither of the participants from this group responded to the online surveys. Each of the 180 questions required participants to identify whether or not the evaluation under consideration contained an element of the *Program Evaluation Standards* (Yarbrough et al., 2011). For each of the thirty standards, there are six yes/no questions or checkpoints. If the participant identified that all six checkpoints were visible in the evaluation it received a score of 6 and that standard was rated as "Excellent". If the participant identified that five of the six checkpoints were visible in the evaluation, the standard received a score of 5 and was rated as "Very Good". If the participant identified that four of the six checkpoints were visible in the evaluation, the standard received a score of 4 and was rated as "Good". If the participant identified that two or three of the six checkpoints were visible in the evaluation the standard was rated as "Fair". If the participant identified that zero or one of the six standards were visible in the evaluation the standard was rated as "Poor". Stufflebeam (2011) arrives at a summative rating for each of the five categories by weighting the score. Each score of "Excellent" is multiplied by 4. Each score of "Very Good" is multiplied by 3. Each score of "Good" is multiplied by 2. Each score of "Fair" is multiplied by 1. The weighted scores are summed and divided by the maximum number of points available for that category to arrive at an overall percentage score, which is then translated back into an overall rating of "Excellent", "Very Good", "Good", "Fair", or "Poor".

In order to arrive at a score for the evaluation based on the evaluator's point of view, I completed the *Checklist,* at the end of the program evaluation. In addition, I hired an expert metaevaluator to review the evaluation and to complete the *Checklist* based on his findings. I did not change any of my ratings after reviewing the metaevaluator's report in order to maintain the integrity of the process. A side by side comparison of the results obtained by each of the individuals or groups who completed the *Checklist* can be found in Appendix K. Descriptive statistics about the metaevaluation scores appear below in Table 9. An in-depth discussion of each evaluation question appears in the following sections.

Table 9 Average Metaevaluation Ratings and Descriptive Statistics

	Standard	Mean	Standard Deviation	Variance
U1	Evaluator Credibility	5.68	.92	.89
U2	Attention to Stakeholders	5.68	.63	.42
U3	Negotiated Purpose	5.36	1.11	1.29
U4	Explicit Values	5.55	.94	.93
U5	Relevant Information	5.41	.98	1.02
U6	Meaningful Processes & Purposes	5.64	.57	.34
U7	Timeliness & Appropriate Communication & Reporting	5.32	.92	1.19
U8	Concern for Consequences & Influence	3.82	1.19	1.49
F1	Project Management	4.95	1.11	1.28
F2	Practical Procedures	5.09	.95	.94
F3	Contextual Viability	4.59	1.19	1.49
F4	Resource Use	3.36	2.25	5.29
P1	Responsiveness & Inclusive Orientation	5.41	.83	.73
P2	Formal Agreements	3.00	1.00	2.00
P3	Human Rights and Respect	4.95	1.4	2.05
P4	Clarity and Fairness	5.18	.78	.63
P5	Transparency and Disclosure	5.14	1.18	1.46
P6	Conflicts of Interest	1.64	.98	1.00
P7	Fiscal Responsibility	1.00	1.00	2.00
A1	Justified Conclusions and Decisions	5.23	1.04	1.14
A2	Valid Information	5.00	1.00	2.00
A3	Reliable Information	3.00	0.00	0.00

A4	Explicit Program and Context Description	4.64	1.67	2.91
A5	Information Management	5.00	1.00	2.00
A6	Sound Design and Analyses	6.00	0.00	0.00
A7	Explicit Evaluation Reasoning	5.05	1.52	2.43
A8	Communication and Reporting	3.50	1.50	4.50
E1	Evaluation Documentation	3.50	0.50	0.50
E2	Internal Metaevaluation	4.00	2.00	8.00
E3	External Metaevaluation	5.50	0.50	0.50

Research Question 2 – Utility

The standard of Utility is designed to measure whether the evaluation is aligned with stakeholder needs in such a way that the results of the evaluation will appropriately serve the information needs of its intended users (Stufflebeam, 2011). The results of the metaevaluation for Utility appear in Table 10. All of the evaluators gave an overall rating of "Very Good" for the evaluation. The lowest marks in this category occurred in the standard of "Concern for Consequences and Influence". The evaluators did not see strong evidence that this evaluation, which used the IEP, put an emphasis on deterring stakeholder sabotage, misuse of findings, or following up after the evaluation to determine how the stakeholders applied the findings. These results should not be interpreted as a conclusive judgment about the utility of the IEP, but rather as a preliminary empirical analysis based on a single case study.

Table 10 Utility Ratings

	Standard	Stake-holders	Meta-Evaluator	Evaluator
U1	Evaluator Credibility	Excellent	Excellent	V. Good
U2	Attention to Stakeholders	Excellent	Excellent	Excellent
U3	Negotiated Purpose	V. Good	Excellent	Excellent
U4	Explicit Values	Excellent	Excellent	Excellent
U5	Relevant Information	V. Good	V. Good	Excellent
U6	Meaningful Processes & Purposes	Excellent	Excellent	Excellent
U7	Timeliness, Appropriate Communication & Reporting	Excellent	Fair	Good
U8	Concern for Consequences & Influence	Good	Fair	Good
	# of Excellent Scores x4	20	20	20
	# of Very Good Scores x 3	6	3	3
	# of Good Scores x 2	2	0	4
	# of Fair Scores x1	0	2	0
Utility Total Score/32 (max possible points available)		88%	78%	84%
Overall Rating for Utility		**Very Good**	**Very Good**	**Very Good**

Research Question 3 – Feasibility

The standard of Feasibility measures the degree to which an evaluation is viable, realistic, contextually sensitive, responsive, prudent, diplomatic, politically viable, efficient, and cost effective (Stufflebeam, 2011). The results of the metaevaluation for Feasibility appear in Table 11. The evaluator gave this standard a rating of "Very Good". The stakeholder group and the metaevaluator gave this standard a rating of "Good". The metaevaluator did not see strong evidence of a formal management plan, provisions for an evaluation staff, or a detailed budget. These are fair assessments of the current evaluation, since it was conducted informally, solely by the evaluator, and at no cost to the client. However, the operationalized version of the IEP contains provisions for these standards to be met in more formal evaluation circumstances. These

results should not be interpreted as a conclusive judgment about the feasibility of the IEP, but rather as a preliminary empirical analysis based on a single case study.

Table 11 Feasibility Ratings

	Standard	Stake-holders	Meta-Evaluator	Evaluator
F1	Project Management	V. Good	Good	Excellent
F2	Practical Procedures	V. Good	V. Good	Excellent
F3	Contextual Viability	V. Good	V. Good	Good
F4	Resource Use	Fair	Good	Good
	# of Excellent Scores x4	0	0	8
	# of Very Good Scores x 3	9	6	0
	# of Good Scores x 2	0	4	4
	# of Fair Scores x1	1	0	0
Feasibility Total Score/16 (max possible points available)		63%	63%	75%
Overall Rating for Feasibility		**Good**	**Good**	**Very Good**

Research Question 4 – Propriety

The standard of Propriety ensures that an evaluation is conducted properly, fairly, legally, ethically, and justly (Stufflebeam, 2011). The results of the metaevaluation for Propriety appear in Table 12. I did not ask the representative stakeholder group to answer all of the questions in this section because they did not possess the expertise required to adequately address all of the standards. However, among the standards that they addressed, they gave ratings of "Very Good" and "Excellent". The evaluator gave this standard an overall rating of "Very Good". The metaevaluator gave this standard an overall rating of "Good". As with the previous standard, the metaevaluator did not see strong evidence of a formal management plan or budget. He also did not see a systematic approach to conflicts of interest. While this is a fair assessment of the current evaluation, the operationalized version of the IEP explicitly addresses conflicts of interest (see Step 7.4 of the IEP). These results should not be interpreted as a conclusive judgment about

the propriety of the IEP, but rather as a preliminary empirical analysis based on a single case study.

Table 12 Propriety Ratings

	Standard	Stake-holders	Meta-Evaluator	Evaluator
P1	Responsive and Inclusive Orientation	Excellent	V. Good	V. Good
P2	Formal Agreements	NA	Fair	Good
P3	Human Rights and Respect	V. Good	Excellent	Excellent
P4	Clarity and Fairness	V. Good	Good	V. Good
P5	Transparency and Disclosure	V. Good	Excellent	Excellent
P6	Conflicts of Interest	NA	Fair	V. Good
P7	Fiscal Responsibility	NA	Poor	Fair
	# of Excellent Scores x4	NA	8	8
	# of Very Good Scores x 3	NA	3	9
	# of Good Scores x 2	NA	2	2
	# of Fair Scores x1	NA	2	1
Propriety Total Score/28 (max possible points available)		NA	54%	71%
Overall Rating for Propriety		**NA**	**Good**	**Very Good**

Research Question 5 – Accuracy

The standard of Accuracy is intended to ensure that an evaluation uses sound theory and reasoning, as well as a solid design, in order to minimize inconsistencies and misconceptions and to produce thoughtful and truthful evaluation findings and conclusions (Stufflebeam, 2011). The results of the metaevaluation for Accuracy appear in Table 13. Again, I did not ask the representative stakeholder group to answer all of the questions in this section because they did not possess the expertise required to adequately address all of the standards. The metaevaluator and the evaluator gave this standard a rating of "Very Good". The metaevaluator's concerns in this area revolved around determining and reporting reliability throughout the evaluation process, as well as a concern about formal agreements concerning dissemination of the evaluation's findings. While this is a fair assessment of the current evaluation, the operationalized version of

the IEP explicitly addresses reliability and validity issues during the evaluation process (Step 6.1). These results should not be interpreted as a conclusive judgment about the accuracy of the IEP, but rather as a preliminary empirical analysis based on a single case study.

Table 13 Accuracy Ratings

	Standard	Stake-holders	Meta-Evaluator	Evaluator
A1	Justified Conclusions and Decisions	Excellent	Excellent	V. Good
A2	Valid Information	NA	Excellent	Good
A3	Reliable Information	NA	Fair	Fair
A4	Explicit Program and Context Descriptions	Excellent	Excellent	Excellent
A5	Information Management	NA	Good	Excellent
A6	Sound Design and Analyses	NA	Excellent	Excellent
A7	Explicit Evaluation Reasoning	Excellent	Good	V. Good
A8	Communicating and Reporting	NA	Fair	V. Good
	# of Excellent Scores x4	NA	16	12
	# of Very Good Scores x 3	NA	0	9
	# of Good Scores x 2	NA	6	2
	# of Fair Scores x1	NA	1	1
Accuracy Total Score/32 (max possible points available)		NA	72%	75%
Overall Rating for Accuracy		**NA**	**Very Good**	**Very Good**

Research Question 6 – Evaluator Accountability

The standard of Evaluator Accountability is "intended to ensure that an evaluation is systematically, thoroughly, and transparently documented and then assessed, both internally and externally" (Stufflebeam, 2011, p. 20). The results for the metaevaluation for Evaluator Accountability appear below in Table 14. The stakeholder participants did not address this standard at all. The metaevaluator gave this standard a rating of "Good" while the evaluator gave this standard a rating of "Very Good". As in previous sections, the metaevaluator's concerns revolved around a lack of formal agreements, budgets, and cost records. These results should not be interpreted as a conclusive judgment about the evaluator accountability of the IEP, but rather as a preliminary empirical analysis based on a single case study.

Table 14 Evaluator Accountability Ratings

	Standard	Stake-holders	Meta-Evaluator	Evaluator
E1	Evaluation Documentation	NA	Fair	Good
E2	Internal Metaevaluation	NA	Fair	Excellent
E3	External Metaevaluation	NA	Excellent	V. Good
	# of Excellent Scores x4	NA	4	4
	# of Very Good Scores x 3	NA	0	3
	# of Good Scores x 2	NA	0	2
	# of Fair Scores x1	NA	2	0
Accountability Total Score/12 (max possible points available)		NA	50%	75%
Overall Rating for Evaluator Accountability		**NA**	**Good**	**Very Good**

Research Question 7 – Overall Effectiveness

Research Question 7 is "how effective is the IEP for evaluating educational programs"? I addressed this question in two ways. First, I calculated an overall score for the current evaluation based on the method advocated by Stufflebeam (2011). Next, I tallied a list of the *Program Evaluation Standards* (Yarbrough et al., 2011) that were identified by Fitzpatrick et al. (2011) as particularly relevant for judging an evaluation design and used Stufflebeam's method to calculate a score for the evaluation design. Using this number, I placed the IEP among the nine models that Stufflebeam & Coryn (2012) evaluated in the most recent edition of their book *Evaluation theory, models, and applications*. This placement represents the summative conclusion of this empirical examination of the IEP.

Stufflebeam's (2011) *Checklist* requires the evaluator to collect and summarize the data for each of the categories and present it in a chart. The chart indicates a percentage score for each category as well as a descriptive term to indicate its overall rating. The results of this metaevaluation are shown below in Table 15 and Figure 6. The stakeholders evaluated two of the

five categories and arrived at a rating of "Excellent" for Utility and "Good" for Feasibility. The metaevaluator rated Utility and Accuracy as "Very Good" and Feasibility, Propriety and Evaluator Accountability as "Good". The evaluator rated all five categories as "Very Good".

Table 15 Bottom Line Results of Summative Metaevaluation

	Stakeholders		Metaevaluator		Evaluator	
	Score	Rating	Score	Rating	Score	Rating
Utility	88%	Excellent	78%	Very Good	84%	Very Good
Feasibility	63%	Good	63%	Good	75%	Very Good
Propriety	NA	NA	54%	Good	71%	Very Good
Accuracy	NA	NA	72%	Very Good	75%	Very Good
Evaluator Accountability	NA	NA	50%	Good	75%	Very Good

Figure 6 Summative Metaevaluation Results
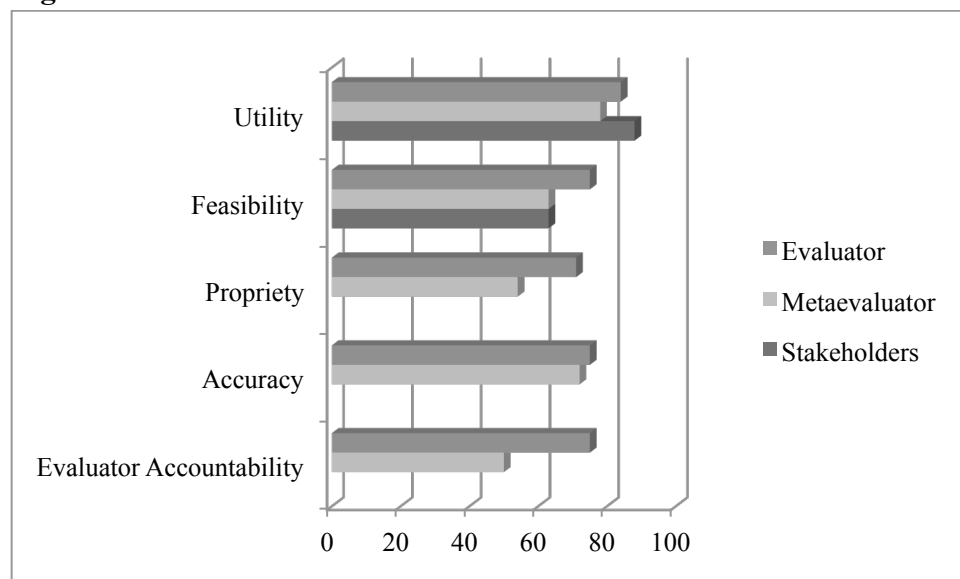
Stufflebeam (2011) proposes two alternative methods to arrive at a summative conclusion about an evaluation. The first method is to average the scores from each category in order to arrive at a score for overall merit. The second method is to calculate an alternative assessment of overall merit by assigning weights to each category based on their relative importance. Stufflebeam recommends assigning a weight of 23% to Utility, 9% to Feasibility, 25% to

Propriety, 34% to Accuracy, and 9% to Evaluator Accountability. Since the representative stakeholders who took part in the evaluation did not address all of the standards, I did not calculate a summative score for them. The metaevaluator's unweighted score was 63% and his weighted score was 66%. Both of these fall into Stufflebeam's "Good" category. The evaluator's unweighted score was 76% and weighted score was 75%. Both of these fall into Stufflebeam's "Very Good" category. These results are summarized in Table 16 below.

Table 16 Assessment of Overall Merit

	Stakeholders		Metaevaluator		Evaluator	
	Score	Rating	Score	Rating	Score	Rating
Overall Merit (Unweighted)	NA	NA	63%	Good	76%	Very Good
Overall Merit: Utility=23% Feasibility=9% Propriety=25% Accuracy=34% Accountability=9%	NA	NA	66%	Good	75%	Very Good

Stufflebeam & Coryn (2012) based their ratings on their combined expertise in metaevaluation and on their experience with actual evaluations conducted using each of the approaches. Since the participants in the current evaluation do not possess the same level of expertise as Stufflebeam and Coryn, I examined the level of interrater reliability among the participants. Wingate (2009) points out that "metaevaluators' interpretation and application of the *Standards* may be mediated by numerous factors that have little to do with the actual quality of the evaluation assessed" (p. 8). Therefore, "reaching reliability is a strong defense against reaching erroneous, invalid conclusions" (p. 8). Fleiss' kappa (1971) is a statistical measure that assesses reliability among more than two raters who have assigned categorical ratings to items.

Fleiss' kappa scores for this evaluation are shown in Table 17. The value of Fleiss' kappa for the 22 participants in the metaevaluation was .604, which Landis & Koch (1977) interpret as indicating substantial agreement among the raters (see Table 6). If the twenty stakeholder participants are combined into one group representing all stakeholders, the value of Fleiss' kappa for the 3 evaluation groups (stakeholders, evaluator, and metaevaluator) increases to .874, which Landis & Koch (1977) interpret as almost perfect agreement among the raters (see Table 6).

Table 17 Fleiss' Kappa

Scoring Method	Fleiss' Kappa	Landis & Koch Interpretation
Counting each participant individually (n=22)	.604	Substantial Agreement
Counting stakeholders as one group (n=3)	.874	Almost Perfect Agreement

Fitzpatrick et al. (2011) state that a metaevaluation of an evaluation design can be conducted by "selecting the *Standards* to be used and applying them to the evaluation design" (p. 373). The standards that they identify as particularly relevant appear in Table 18, along with the scores attributed to them by each of the metaevaluating groups/individuals. The score for the stakeholder group is calculated slightly differently than that of the evaluator and metaevaluator. Since the stakeholders did not address as many standards as the other two groups, the maximum number of points available in their column was 56. The maximum number of points available to the other two groups was 72. The scores were calculated accordingly. Since this scoring method focuses on the standards that are specifically relevant to the design of the evaluation, it returns a result that is particularly interesting as part of the evaluation of the model. In fact, the overall score for the evaluation increases significantly by using this method. The evaluator and metaevaluator arrived at a rating of "Very Good" for the IEP. The stakeholders arrived at a rating of "Excellent" for the IEP. The increase in rating using this method of calculation is likely

due to the fact that this method does not include some of the more procedural elements such as

budgeting, implementation of formal agreements, or metaevaluation. Instead, it focuses on more

conceptual elements such as negotiated purpose, inclusiveness, and fairness.

**Table 18 Design Metaevaluation per Fitzpatrick et al. (2011)**

	Standard	Stake holders	Meta Evaluator	Eval- uator
A6	Sound Design and Analyses	NA	Excellent	Excellent
U3	Negotiated Purpose	Excellent	Excellent	Excellent
P1	Responsiveness and Inclusive Orientation	Excellent	V. Good	V. Good
P3	Human Rights and Respect	Excellent	Excellent	Excellent
P4	Clarity and Fairness	Excellent	Good	V. Good
P5	Transparency and Disclosure	Excellent	Excellent	Excellent
A1	Justified Conclusions and Decisions	Excellent	Excellent	V. Good
A2	Valid Information	NA	Excellent	Good
A3	Reliable Information	NA	Fair	Fair
A4	Explicit Program and Context Descriptions	Excellent	Excellent	Excellent
A5	Information Management	NA	Good	Excellent
U2	Attention to Stakeholders	Excellent	Excellent	Excellent
U4	Explicit Values	Excellent	Excellent	Excellent
U6	Meaningful Processes and Purposes	Excellent	Excellent	Excellent
F1	Project Management	Excellent	Good	Excellent
F2	Practical Procedures	Excellent	V. Good	Excellent
F3	Contextual Viability	Excellent	V. Good	Good
F4	Resource Use	Excellent	Good	Good
	# of Excellent Scores x 4	56	40	44
	# of Very Good Scores x 3	0	9	9
	# of Good Scores x 2	0	8	6
	# of Fair Scores x1	0	1	1
	Accountability Total Score/ 56 or 72 (max possible points available)	100%	81%	83%
	Overall Rating	**Excellent**	**Very Good**	**Very Good**

Stufflebeam & Coryn (2012) used the *Checklist* to calculate a score for nine of the

evaluation approaches that they deem to be the best for 21st century evaluations. These include

Patton's (1986) Utilization Method, Stake's (1975) Responsive/Client-Centered approach and

Case Study Method, Guba and Lincoln's (2004) Constructivist Approach, Scriven's (1991)

Consumer Oriented method, Tyler's (1942) Objectives Based approach, Campbell & Stanley's (1966) Experimental and Quasi-experimental design, Brinkerhoff's (2003) Success Case Method, and Stufflebeam's own (1983) CIPP Model. Based on their analysis, they judged that only three of the approaches earned an overall rating of "Very Good", including Stufflebeam's (1983) CIPP model, which earned the highest rating. Six approaches earned an overall rating of "Good".

This study was conducted using an operationalized version of the IEP that is keyed to Stufflebeam's (2011) *Checklist.* Twenty-two people participated in the metaevaluation, and had a high rate of interrater reliability. The evaluation design earned a rating of "Very Good" by the evaluator and the professional metaevaluator. The participant stakeholders rated the design as "Excellent". The placement of the IEP among the models analyzed by Stufflebeam and Coryn (2012) is shown below in Figure 7. Of course, this placement is not definitive because it is based on a single case study. However, it provides an initial illustration of the IEP based on empirical evidence.

Figure 7 Evaluation Model Ratings
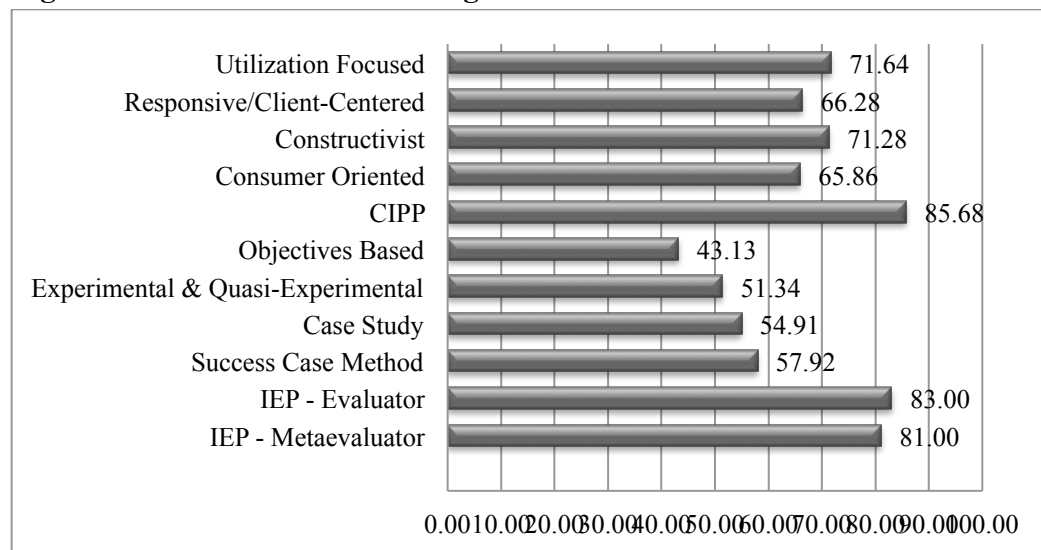
Summary

The purpose of this study is to identify where the Guerra-Lopez IEP fits within the body of evaluation theory, to assess the effectiveness of the IEP as a tool to evaluate educational programs, and to empirically examine the IEP from multiple perspectives, primarily using Stufflebeam's (2011) *Metaevaluation Checklist*. Through the review of the literature and a comparison of the characteristics of the IEP with those of the leading models in the field of evaluation, it becomes clear that the IEP shares the most common elements with models that focus on how the evaluation results will be used by the stakeholders. Fitzpatrick et al. (2011) refer to these types of evaluations as "decision oriented". Christie and Alkin (2012) classify these models on the "Use" branch of their metaphorical tree. Stufflebeam (2011) categorizes these models as "Eclectic".

There is no generally accepted method for empirically evaluating an evaluation model. A variety of scholars have proposed frameworks for doing so, and I used several of them to begin to draw conclusions about the IEP. Fitzpatrick et al. (2011) and Miller (2010) proposed methods that revolve around the use of an operationalized version of the model. Therefore, I developed an operationalized version of the IEP as part of this study. Stufflebeam (2011) proposes the use of his *Program Evaluation Metaevaluation Checklist* to evaluate models. Since this type of metaevaluation is based on the review of an actual evaluation, I conducted an evaluation of a 1:1 technology program. I asked the stakeholders who participated in the evaluation to complete Stufflebeam's (2011) *Checklist*. In addition, I completed it as the evaluator, and I hired a professional metaevaluator to review the evaluation and complete the *Checklist*. The results of these metaevaluations culminated in an overall rating for this evaluation ranging from "Good" to

"Very Good". Fitzpatrick et al. (2011) identify certain standards within the *Checklist* that are particularly relevant to evaluation design. When I examined the results of the evaluation using these standards, the overall rating for the evaluation design ranged from "Very Good" to Excellent". In the next chapter, I discuss the meaning of these findings and the implications for the field of Performance Improvement.

CHAPTER 5 Discussion

Introduction

The purposes of this qualitative single case study included an examination of the theoretical underpinnings of the Guerra-Lopez (2007a) Impact Evaluation Process (IEP) as well as an evaluation of its effectiveness in practice. By addressing both of these issues, this research project clarifies the connection between theory and practice. The study examined seven research questions:

1: Where does the Guerra-Lopez IEP fit into the body of evaluation research?

2: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Utility?

3: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Feasibility?

4: To what extent does the IEP meet the Joint Committee Program Evaluation requirements for Propriety?

5: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Accuracy?

6: To what extent does the IEP meet the Joint Committee Program Evaluation Standard requirements for Evaluator Accountability?

7: How effective is the IEP for evaluating educational programs?

This chapter includes a brief summary of the study and its findings. It discusses the conclusions that can be drawn from the study as well as the limitations. It makes

recommendations for future research, and concludes with a discussion of the implications of this study for the field of Performance Improvement.

Summary and Findings

There is no generally accepted process for arriving at a final judgment of the usefulness, appropriateness, effectiveness, reliability, and validity of an evaluation model. Several people have proposed methods. Stufflebeam and Coryn (2012) use Stufflebeam's (2011) *Checklist* to make judgments about popular evaluation models based on their professional expertise and their review of actual evaluations conducted using the models. Miller (2010) proposes a second method to empirically evaluate how theory informs practice and whether particular theories yield superior evaluations. Miller's framework consists of five criteria: operational specificity, range of application, feasibility in practice, discernible impact, and reproducibility. Miller believes that, in order for a theory to be useful, it must offer specific guidance for practice. This includes providing procedural guidelines for conducting the evaluation. Fitzpatrick et al. (2011) propose a third method to conduct metaevaluations of evaluation designs. They also believe that the existence of an operationalized version of the design is critical. They propose that the evaluator select certain *Program Evaluation Standards* (Yarbrough et al., 2011) to be used and "apply them to the design at various stages of the evaluation" (p. 373). They provide a list of 18 appropriate *Standards* to consider. Next, they suggest that the evaluator needs to use this information to make a judgment about whether, on balance, the design "seems to achieve its purposes at an acceptable level of quality" (p. 373). In this study, I used a combination of these approaches to begin to build a body of evidence about the effectiveness of the Guerra-Lopez IEP.

First, through a study of three evaluation theory classification schemes, I was able to conceptually identify where the IEP fits among other common evaluation models. Clearly, the IEP most closely resembles and shares many common elements with Patton's (1986) Utilization-focused model. Both models are primarily concerned with impact, rely on the input of representative stakeholders to focus the evaluation, and use mixed methods to achieve the desired results. The IEP stresses the impact on organizational goals, clients and society while the Utilization approach focuses more on ensuring that the results are *used* by stakeholders to fulfill their stated objectives.

Next, it was apparent that a judgment about the model could not be reached without evidence based on its application in the real world. While other evaluations have been conducted using the IEP (Guerra-Lopez & Toker, 2012), I did not have access to sufficient information from them to make sound judgments. Therefore, I planned and executed an evaluation of an educational program using the IEP. Before and during the technology program evaluation, I developed an operationalized version of the IEP. I did this in order to meet Fitzpatrick et al. (2011) and Miller's (2010) requirement that empirical examinations of evaluation models include operational specificity.

The Joint Committee on Standards for Educational Evaluation (2010) has identified a set of 30 *Program Evaluation Standards* for the evaluation of educational programs. Stufflebeam (2011) has further elaborated on these standards by creating a list of 180 checkpoints that relate to the thirty standards. As part of the operationalization of the IEP, I associated each of the 180 points to a step within the model (see Table 7). During the evaluation of the technology program, and at its conclusion, I asked the participants to judge the evaluation based on Stufflebeam's (2011) *Checklist.* At the conclusion of the evaluation, I also completed the *Checklist* in order to

arrive at my own conclusion about the evaluation. In addition, I hired a professional metaevaluator to conduct a metaevaluation based on the *Checklist*. While I cannot claim that a final judgment about the model can be made through a single case study such as this, I can use the results from the three groups who participated in the metaevaluation to begin to build an empirical body of evidence about the model's effectiveness. The three groups had a high level of agreement that the evaluation addressed most of the checkpoints that are tied to the *Program Evaluation Standards*. For the standard of Utility, all three groups gave an overall rating of "Very Good". For the standard of Feasibility, the stakeholders and the metaevaluator gave a rating of "Good" and the evaluator gave a rating of "Very Good". I did not ask the stakeholders to rate the remaining standards of Propriety, Accuracy, and Evaluator Accountability because these standards required a certain level of expertise, or at least familiarity, with technical terms within the field of evaluation. The metaevaluator and evaluator agreed on a rating of "Very Good" for the standard of Accuracy. The metaevaluator gave a rating of "Good" for the standards of Propriety and Evaluator Accountability, while the evaluator rated both of these standards as "Very Good". These consistently high ratings among the three metaevaluating groups suggest that the use of the IEP to conduct an evaluation will lead to effective evaluations that are tied directly to the *Program Evaluation Standards*.

The IEP received the highest ratings by all three metaevaluating groups in the area of Utility. This standard measures whether an evaluation is aligned with stakeholder needs in such a way that the results will effectively provide users with appropriate and necessary information. This is a particularly crucial standard in an impact evaluation, where the emphasis is on meeting stakeholder and decision-makers' needs. The high scores in this standard suggest that the IEP is effective for conducting impact evaluations.

The IEP received the next highest ratings in the standard of Accuracy. This standard measures whether or not an evaluation uses sound theory and reasoning, as well as solid design, in order to produce thoughtful and truthful evaluation findings and conclusions (Stufflebeam, 2011). This is a crucial standard for measuring the effectiveness of an evaluation design or model, and the high marks in this standard indicate that the IEP is based on a sound theoretical foundation, which will ultimately lead to reliable findings and conclusions.

There were several characteristics of this evaluation that undoubtedly affected the overall rating by each of the evaluating groups/individuals. Since the evaluation was conducted as part of this research study, there was no cost to the client. Therefore, all metaevaluation checkpoints that addressed fiscal responsibility were not checked. In addition, since the client and all of the stakeholders viewed the evaluation as a welcomed endeavor, I did not have to manage conflicts of interest or potential stakeholder sabotage. There was, in fact, uniform enthusiasm for the project. Therefore, I was unable to provide evidence that these were addressed as part of the evaluation.

By using Stufflebeam's method for arriving at a bottom line result for a summative metaevaluation, the metaevaluator gave the evaluation an overall rating of 66% which is translated as "Good" while the evaluator gave it an overall rating of 75% which is translated as "Very Good". Fitzpatrick et al. (2011) recommend a more specific method of arriving at a summative score for an evaluation design by looking at 18 of the Joint Committee's (2011) thirty *Standards* that are particularly relevant for examining the model behind the evaluation. These include the following: sound design and analysis, negotiated purpose, responsiveness and inclusive orientation, human rights and respect, clarity and fairness, transparency and disclosure, justified conclusions and decisions, valid information, reliable information, explicit program and

context descriptions, information management, attention to stakeholders, explicit values, meaningful processes and purposes, project management, practical procedures, contextual viability, and resource use. They stress that no design is perfect, and that the judgment is whether, "on balance, after summarizing judgments across scales, the evaluation seems to achieve its purposes at an acceptable level of quality" (p. 373).

Ultimately, this method produced the results that I believe are the most valuable for this study. While the Stufflebeam method that I used above provided interesting results, the single case study was a significantly limiting factor for this approach. There were several unique elements to the evaluation that contributed to low scores on several standards. This limitation would be mitigated by using the *Checklist* to evaluate several evaluations conducted using the model. However, the Fitzpatrick et al. (2011) method allowed me to separate the results of the evaluation for these 18 standards and to filter out some of the standards that did not apply to this evaluation. By using this method, the evaluator and metaevaluator arrived at overall scores of 81% and 83% respectively for this evaluation design.

In order to put these numbers into context, I compared them to the scores that Stufflebeam and Coryn (2012) arrived at for nine of the most effective evaluation models in use today. This comparison provides a basis for future metaevaluation. Stufflebeam and Coryn used a combination of actual evaluations and their own expertise to make a summative judgment about the models. Their scores for these nine models ranged from 43-86%. The highest score that Stufflebeam and Coryn (2012) awarded in their study was 85.68% for Stufflebeam's own CIPP (1983) model. The next highest score was for Patton's (1986) Utilization-focused model (71.64%). During my examination of the placement of the IEP among other models within the field of evaluation, I concluded that the IEP shares many common conceptual elements with

Patton's (1986) model and placed it near Patton and Stufflebeam on the evaluation theory "tree" described by Christie and Alkin (2012) (see Figure 4). Since the three models share key conceptual similarities while retaining unique theoretical elements, it is not surprising that they received similar high scores using Stufflebeam's (2011) *Checklist*. The fact that the three models received similar scores further triangulates the results obtained in this study and indicates that the use of the model will lead to an effective evaluation that is based on a strong theoretical foundation.

Miller's (2010) recommended framework for empirically evaluating evaluation models includes five criteria: operational specificity, range of application, discernible impact, reproducibility, and feasibility in practice. This research study addressed several of these criteria. In particular, one of the key elements of the study was the development of an operationalized version of the IEP. The operationalized model provides procedural guidelines regarding key issues such as what questions to ask, whom to include in the process, what methods to use, how the values that undergird the model are best enacted, and what role the evaluator will play. The operationalized IEP includes each of the seven steps identified by Guerra-Lopez (2007a) as well as activities that need to occur prior to the evaluation and activities that occur iteratively throughout the process. In order to make a theoretical connection between the IEP, the *Program Evaluation Standards (*Yarbrough et al., 2011), and Stufflebeam's (2011) *Checklist* I tied the steps in the IEP to each of the 180 checkpoints.

Because of the nature of a single case study, I did not address Miller's criterion of range of application. This criterion seeks to consider the described limits of a theory's application, and its adaptability across a wide range of conditions. The criterion of discernible impact is concerned with an examination of whether the use of a theory really leads to the impact that is

expected. Theoretically speaking, there should be discernible benefits "because of, due to, and linked to" (Miller, 2010, p. 395) the evaluation approach. This study did not specifically address discernible impact. However, I carefully included the theoretical "fingerprint" of the IEP in the operationalized version of the model so that the elements of the model that are unique are clearly visible. The criterion of reproducibility calls for an examination of whether an approach "can achieve its intended outcomes in diverse evaluators' hands" (p. 395). Although this study does not specifically address this issue, the use of the operationalized version of the IEP will undoubtedly increase the ability of diverse evaluators to successfully conduct appropriate evaluations.

The criterion of feasibility in practice was addressed by the use of Stufflebeam's Checklist. The standard of feasibility analyzes four main elements: project management, practical procedures, contextual viability, and resource use. The stakeholders gave an overall rating of "Good' for feasibility. The metaevaluator rated project management and resource use as "Good" and practical procedures and contextual viability as "Very Good". The evaluator rated project management and practical procedures as "Excellent" and contextual viability and resource use as "Good". Overall, the three groups gave a summative rating for feasibility that ranged from "Good" to "Very Good".

Overall, there are two points that provide the most compelling evidence about the IEP. First, the fact that the model received particularly strong results in Accuracy and Utility indicate that it is an effective design that is based on a strong theoretical foundation, and that it will lead to evaluations that provide users with information upon which they can make sound decisions. Second, the fact that the design received scores that are similar to the scores received by Stufflebeam's CIPP model (1983) and Patton's Utilization-focused (1986) model suggests that

the IEP can be placed among the models that have been identified as the most effective models in use today.

Limitations

This study has several limitations. A primary concern and a potential weakness of the study is the question of how a single evaluation can effectively allow us to render a judgment about an evaluation model. At the beginning of this project, my goal was to imitate Stufflebeam and Coryn (2012) by rendering a definitive judgment about the IEP based on my "experience in seeing and assessing how (it) worked in practice" (p. 443). I planned to make up for my relative inexperience by triangulating my results with those of the evaluation participants and the professional metaevaluator. However, during the course of the project I realized that the particular idiosyncrasies of this evaluation would affect the overall rating by the metaevaluation participants and therefore preclude definitive conclusions about the model behind the evaluation process. However, all is not lost. The goal of the empirical evaluation of an evaluation model is to determine its "ultimate feasibility and merit" (Miller, 2010, p. 391). The *Program Evaluation Standards* (Yarbrough et al., 2011) are widely accepted as "principles mutually agreed to by people engaged in the professional practice of evaluation, that, if met, will enhance the quality and fairness of an evaluation" (Joint Committee, 2010, p.3). By operationalizing the IEP and aligning it not only with the 30 *Standards*, but also with Stufflebeam's (2011) 180 checkpoints, and then by conducting an evaluation and a metaevaluation using the operationalized version, I am providing evidence that the model contains the principles of the Standards. While the exact score earned by individual evaluations based on the *Checklist* will undoubtedly vary, there is reasonable certainty that an evaluation conducted using the IEP will earn a satisfactory score.

Ultimately, based on these realities and realizations, I changed my expectations for my conclusions. Instead of claiming to make a definitive judgment about the effectiveness of the IEP, I am only claiming to make preliminary judgments that need to be verified with additional research.

A second limitation of the study is the inherent potential for bias that exists in the metaevaluation process. As the evaluator and as a doctoral candidate, I had the ability to affect and be affected by the outcome of the evaluation. I controlled the information that was delivered to the stakeholder participants and the metaevaluator. In addition, I will be affected by the judgment of the merit of the research study. As a result, in a sense I became a stakeholder in the process. This is somewhat unavoidable because evaluation is necessarily a subjective process. However, the purpose of metaevaluation is to summarize judgment across scales and to determine whether the model achieves its intended purpose at an acceptable level (Fitzpatrick et al., 2011).

This study had several inherent challenges from a statistical standpoint. First, I divided the Stufflebeam (2011) *Checklist* into two parts and created an online survey for each part. I asked the representative stakeholders to take the first survey in mid –June 2012. The questions on this portion of the survey revolved around the evaluation process. I asked the stakeholders to take the second survey in mid-September 2012. The questions on this portion of the survey revolved around the evaluation findings and report. Quantitative researchers strongly discourage dividing surveys into multiple parts because of the possibility of resulting bias (Kish, 1965). An additional concern revolves around Stufflebeam's (2011) *Checklist,* which is difficult to analyze statistically because it calculates ratings based on unequal intervals, and switches between

nominal and ordinal data. However, since a statistical analysis of the *Checklist* survey results is not central to the study, these concerns are not catastrophic.

Future Research

This study represents the first systematic attempt to empirically evaluate the IEP, but additional research is needed. In order to strengthen the results of this single case study, it would be valuable to do full metaevaluations of additional evaluations that are conducted using the IEP. The current study rated the IEP highest in Utility and Accuracy, and rated it lowest in Evaluator Accountability and Propriety. However, these results are partially attributable to the unique conditions under which the evaluation occurred. Through repeated use of Stufflebeam's *Checklist* to conduct metaevaluations, stronger patterns of strength and weakness are likely to emerge.

Miller (2010) recommends a framework for empirically evaluating how theory informs practice. In this study, I was able to address several of Miller's criteria for guiding research on evaluating the relationship between theory and practice. However, I was unable to significantly address the criteria of range of application and discernible impact. Miller (2010) proposed this framework as a way to address the lack of guidance that exists to help researchers empirically examine evaluation theory. However, the framework lacks details about *how* to examine each criterion. I was able to operationalize the model under consideration and address the issue of feasibility in practice using the *Checklist*, but the other criteria were out of my reach. Miller's (2010) framework requires the operationalization of evaluation theory. In turn, the field would benefit from the operationalization of Miller's framework.

Implications for Performance Improvement

One of the overarching goals of this research study is to contribute to the empirical research base of the field of Performance Improvement. While the field continues to grow and develop, it faces several ongoing challenges. Its eclectic elements, which draw from several applied fields such as organizational development, may contribute to a lack of clarity (Pershing et al., 2008). The field faces occasional criticism that the empirical foundations have not kept pace with practice (Stolovitch, 2000; Sugrue & Stolovitch, 2000).

Evaluation is a core element of the Performance Improvement process as well as a field of study in its own right. The field of evaluation is a mature professional discipline with common terminology and professional standards. The American Evaluation Association has agreed upon *Guiding Principles for Evaluators* (Fitzpatrick et al., 2010). The Joint Committee on Standards for Educational Evaluation publishes a set of *Standards for Program Evaluation (*Yarbrough et al., 2011) which is designed to help consumers and evaluators judge the quality of evaluations. Professionals within the field of evaluation subscribe to a wide variety of prescriptive theories. There are several theoretical frameworks that evaluators use to categorize these theories.

The Guerra-Lopez (2007a) Impact Evaluation Process (IEP) is a relatively new model within the field of Performance Improvement. It is heavily influenced by Kaufman's (2006) work on needs assessment. While Guerra-Lopez (2008) states that the IEP has been used in a variety of settings, including education, business, and healthcare, it will benefit by additional empirical examination. As part of the review of the literature in the field, I identified that the IEP is most closely aligned with evaluation theories that focus on how evaluation results will be used, and what decisions will be made as a result of the evaluation.

The key contribution of this study is to provide a bridge between the theory and practice of evaluation. Evaluation theory is generally prescriptive, but frequently vague when it comes to

specific steps that should be followed in order to get strong, reliable results. This study addressed this concern by creating an operationalized version of the IEP that is tied directly and specifically to Stufflebeam's (2011) *Metaevaluation Checklist* and to the *Program Evaluation Standards* (2010). This version of the model clearly highlights the IEP's theoretical fingerprint, while making the process of conducting an evaluation accessible to experts and novices alike.

Needs assessment and evaluation are the "bookends" of the field of Performance Improvement. Since the IEP was developed with strong theoretical ties to Kaufman's (2006) work on needs assessment, it allows performance improvement practitioners to work through the performance improvement process seamlessly. It allows even novice practitioners to conduct an evaluation that is grounded in solid theory and professionally accepted standards of excellence. The results of this study indicate that the IEP is particularly strong in the area of Utility. This means that it is focused on serving the needs of stakeholders and decision makers by providing useful information that will allow them to make decisions that will affect positive change within their organization.

The implications of this study for practitioners in the field of evaluation are similar. The operationalized version of the IEP is easy to follow and is clearly tied to the *Program Evaluation Standards*. As the climate in industry is increasingly geared toward data driven decision making and quantifiable results, the IEP provides evaluators with the means to conduct evaluations that not only examine outcomes but also help to empirically assess the changes that can be attributed to a program. The IEP's concern with outcomes at the societal, organizational, and operational levels is its most unique characteristic, and the element that makes it unique among evaluation models. By focusing on each of these levels, evaluators can ensure that they provide

comprehensive evaluations that allow stakeholders to make effective decisions and implement changes that will lead to continuous organizational improvement.

This study contributes to the field of evaluation because the metaevaluation that was conducted as part of the study was, to my knowledge, the first to use Stufflebeam's (2011) newly revised *Checklist*. In addition, it is among the first studies to refer to the newly revised *Program Evaluation Standards* (Yarbrough et al., 2011). The study answers the ongoing call from within the field of evaluation for additional metaevaluation in order to continue to gather data that will help to identify the mechanisms that ensure that evaluations achieve their intended purpose (Mark & Henry, 2004).

Summary

There is no one perfect model or design for conducting a program evaluation. Evaluation theorists base their models on their own theoretical perspective and professional values. The Guerra-Lopez (2007a) Impact Evaluation Process emphasizes the idea of searching for relevant, reliable, and valid data to show that a program is helping an organization to reach its desired results at the strategic, tactical, and operational levels. The purpose of this study was to empirically evaluate the IEP in order to make a preliminary judgment about its effectiveness.

This study consisted of an evaluation that was conducted using the IEP and a metaevaluation of that evaluation. As part of the evaluation, an operationalized version of the model was developed. The metaevaluation was based on the Joint Committee's *Program Evaluation Standards* and Stufflebeam's *Program Evaluations Metaevaluation Checklist*. The Stufflebeam *Checklist* leads users ultimately to assign a rating that ranges from "Poor" to "Excellent" for the evaluation under consideration. Since this metaevaluation was particularly

interested in examining the design of the evaluation, it went one step further by selecting the *Program Evaluation Standards* that are focused on evaluation design elements. The score that the evaluator and a professional metaevaluator arrived at for this evaluation design was "Very Good", which ranks it among the top models in the field. The 20 participant stakeholders in the evaluation rated it even higher, as "Excellent". A calculation of reliability among the 22 metaevaluation participants indicates a high level of interrater reliability. The purpose of metaevaluation is to determine "whether, on balance, after summarizing judgments across scales, the evaluation seems to achieve its purposes at an acceptable level of quality" (Fitzpatrick et al., 2011, p. 373). As such, the results of this study indicate that the IEP is clearly tied to the professional standards within the field of evaluation, and achieves its purposes at an acceptable level of quality.

APPENDIX A – APPROVAL TO USE STUFFLEBEAM PROGRAM EVALUATION METAEVALUATION CHECKLIST

WESTERN MICHIGAN UNIVERSITY

Interdisciplinary Ph.D. in Evaluation

February 22, 2012

Dear Mrs. Blake:

This letter is to grant you permission to use the following for your doctoral dissertation research: (1) the checklist based on the 2011 *The Program Evaluation Standards*; (2) an Excel macro that automates the ratings from the checklist; and (3) a draft chapter from the forthcoming book in which myself and Dr. Daniel L. Stufflebeam metaevaluate several evaluation approaches using the checklist. In return, we ask that you do not distribute these materials publically and that you cite them as Stufflebeam, D. L., & Coryn, C. L. S. (2012). *Evaluation theory, models, & applications* (2nd ed.). San Francisco, CA: Jossey-Bass.

Please do not hesitate to contact me if I can be of any further assistance.

Sincerely,

Chris L. S. Coryn, Ph.D.
Assistant Professor of Evaluation, Measurement, and Research
Director of the Interdisciplinary Ph.D. in Evaluation
The Evaluation Center
Western Michigan University
1903 West Michigan Avenue
Kalamazoo, MI 49008-5237
Telephone: 269-387-5906
E-mail: chris.coryn@wmich.edu

Name	Name of Approach	Branch of Theory Tree (2012)	Stufflebeam Categorization	Fitzpatrick, Sanders & Worthen Categorization	Description
Alkin	User-Oriented Evaluation (1991)	Use	Improvement and Accountability Approach	Decision Oriented	- Rejects notion of evaluators as valuing agents - Works with users at outset to establish value systems for judging outcome data - Focus is on identified potential use
Boruch, McSweeney & Soderstrom	Randomized Field Experiment Approach (1978)	Methods	Questions and Methods	NA	- Randomized field experiments are most effective way to obtain least equivocal estimate of social program effects
Brinkerhoff	Success Case Method (2003)	NA	Questions and Methods	NA	
Campbell & Stanley	Experimental and Quasi-Experimental Design (1966)	Methods	Questions and Methods	NA	- Heart & Soul of Methods branch - Campbell is best known for work on elimination of bias in conduct of research in field settings
Chelimsky	Evaluation in Democratic Society (1995)	Use	Improvement and Accountability Approach	NA	"telling the truth to people who may not want to hear it is, after all, the chief purpose of evaluation" (1995) - Evaluation of public programs & policies is fundamental to democratic government - Work focuses on large groups and governmental bodies
Chen & Rossi	Theory Driven Evaluations (1989)	Methods	Questions and Methods	Program Oriented	- Recognize dominance of

					believes it should be supplemented by social science theory to identify additional areas to investigate - Concerned with identifying secondary effects and unintended consequences of programs - Most influential developer of theory driven evaluation
Cook & Campbell	Quasi-Experimental Design (1979)	Methods	Questions and Methods	NA	- Credited with developing field of study related to quasi-experimental design - Concerned with contextual factors of evaluation and ways these factors can affect an evaluation - Focused on using several different designs & methods to properly conduct an evaluation - One of first methods-driven evaluators to recognize importance of involving stakeholders
Cousins & Whitmore	Participatory Evaluation (1998)	Use		Participant Oriented	- In order to heighten possibility of utilization, must have structured, continued, and active participation of intended users - Utilization is best accomplished as part of organizational development - Primary users and evaluators are recognized as collaborators - Preference for research methods
Cronbach & Associates	Reform of Program Evaluation (1980)	Methods	Questions and Methods	Program Oriented	- Sees evaluator's role as providing people with information that they may consider when forming their

					own judgment - Views evaluation as integral part of policy research - Methodological contributions include Cronbach's Alpha, generalizability theory and notions about construct validity - Chiefly concerned with generalizability
Eisner	Educational Connoisseurship (1976)	Valuing	Questions and Methods	Consumer and Expertise Oriented	- Evaluation is making value judgments about quality of object, situation, or process -Focused on twin notions of connoisseurship and criticism -Connoisseurship is to have knowledge about subject and ability to differentiate subtleties - Criticism is making experience public through description, expectation, and background knowledge - Almost exclusively uses qualitative methods
Fetterman	Empowerment Evaluation (1994)	Use	Social Agenda and Advocacy	Participant Oriented	- Proposes to empower most marginalized, oppressed to control their own destiny by use of results of study - Program participants essentially manage their own evaluation - Evaluator coaches or teaches user to conduct evaluation
Greene	Value-Engaged Approach	Valuing		NA	- Includes 3 criteria of deliberative,

	(2005)				democratic evaluation: inclusion, dialogue, and deliberation. - Also stresses stakeholder involvement, which resembles participatory eval approaches. - Emphasizes mixed methods designs and fieldwork
Guba & Lincoln	Naturalistic Evaluation (1981) Fourth Generation Evaluation (2004)	Valuing	Questions and Methods	Participant Oriented	- Stakeholders are primary individuals involved in placing value - There are multiple realities to be evaluated based on perceptions & interpretations of individuals involved - 4th Generation Evaluation is combination of responsive focusing and constructivist methodology aimed at developing consensus among stakeholders
Guerra-Lopez	**Impact Evaluation Process (2007a)**	**Use**	**Eclectic**	**Decision Oriented**	**- Identify stakeholders and expectations** **- Determine key decisions and objectives** **- Derive measurable indicators** **- Identify data sources** **- Select data collection instruments** **- Select data analysis tools** **- Communicate results and recommendations**
Henry, Julnes, and	Emergent Realist Evaluation (1998)	Methods		NA	- Methodology is core of approach - Social betterment is ultimate

Mark					objective of evaluation - Leans toward Values branch of tree - gives priority to the study of generative mechanisms - attentive to multiple levels of analysis - mixed methods appropriate -Evaluation as a tool for social & political change within democracies
House	Evaluation and Social Justice (1991)	Valuing		Program Oriented	- Purpose of evaluation is to provide information to decision makers so they can determine allocation of vital resources - Evaluation is never values neutral, must lean towards social justice by addressing needs of powerless - Don't define value in terms of good or bad, but in terms of right/just/fair - Ontologically and epistemologically work is grounded in constructivist thinking, but methodologically leans toward quantitative.
House & Howe	Deliberative Democratic Model (1999)	NA	Social Agenda and Advocacy	Participant Oriented	- Evaluators should accept authority but not power
King & Stevahn	Interactive Evaluation Practice (2005)	Use		NA	- Designing and implementing evaluations in collaborative manner with stakeholders for purpose of increasing likelihood that information generated from the

					evaluation will be used - Emphasizes participation, capacity and interpersonal factors
Kirkpatrick	Four Levels of Evaluation (1994)	NA		NA	
Levin	Cost-Benefit Analysis (2005)	Valuing	Questions and Methods	NA	- Focus on development, promotion, and use of cost analysis as means for drawing evaluative conclusions - An array of economics-based strategies to determine program costs before and during implementation
Mertens	Inclusive/Transformative Model (2001) Emancipatory Evaluation (1999)	Valuing	Social Agenda and Advocacy *	Participant Oriented	- Emphasis on diversity and inclusion of diverse groups - Evaluator's primary role is to include marginalized groups, not to act as decision maker
Patton	Utilization-focused (1986)	Use	Eclectic	Decision Oriented	- Not primarily focused on decision makers needs but on emphasizing procedures that would enhance use of evaluation to broader spectrum of stakeholders. - Evaluator should seek out individuals who will use the evaluation – intended primary users - Users must commit to intended focus of evaluation - Users should be involved in methods, design, and measurement - Users should be engaged in interpreting findings and making

					judgments - Users should be involved in making decisions about further dissemination
Phillips	Return on Investment Methodology (1997)	NA		NA	
Preskill & Torres	Transformational Learning (2001)	Use		NA	- Focus on organizational learning and development - Substantial evaluation utilization occurs during evaluation process and is tool for transformative learning - Evaluators should provide more than technical expertise to conduct evaluation to allow reflection and creation of dialogue which facilitates transformational learning

Provus	Discrepancy Evaluation Model (1971)	Use (2004) Removed from subsequent versions	Questions and Methods	Program Oriented	- Evaluation is continuous process designed to assist program administrators
					- Four developmental stages:
					  (1) definition – specify goals, processes, and resources
					  (2) installation – identify discrepancies in implementation of program
					  (3) process – determine extent of attainment of enabling objectives
					  (4) product – determine extent of attainment of terminal objectives
					- Final stage = Cost benefit analysis

Rossi,	Theory Driven Evaluation	Methods		NA	- Foundational part of theory driven

137

Freeman & Lipsey	(1999)				evaluation - Theory driven evaluation involves creation of detailed program theory. This is then used to guide evaluation – helps to reconcile internal and external validity
Scriven	Goal Free Evaluation (1972) Consumer Oriented (1991)	Valuing	Improvement and Accountability Approach	Program Oriented; Consumer and Expertise Oriented	- Major contribution is insistence on role of evaluator in making judgments - Evaluation is science of valuing - It is not necessary to explain why a program works to determine its value - Evaluator assumes responsibility for determining which program outcomes to examine/reject. - Use objectives of program as starting point - Identified distinction between formative and summative evaluation
Stake	Case Study Method Responsive Evaluation (1975)	Valuing	Social Agenda and Advocacy	Participant Oriented	- Difficult to categorize - Essential components include belief that knowledge is bound by context – no true value to anything - Stakeholder perspectives are integral elements in evaluation - Case studies are best method for representing beliefs and values of stakeholders and of reporting evaluation results
Stufflebeam	CIPP (1983)	Use	Improvement	Decision	- Four types of evaluation: Context,

138

				Oriented	Input, Process, Product
			and Accountability Approach		- Context evaluation involves identifying needs to decide on program objectives - Input evaluation lead to decisions on strategies and designs - Process evaluation consists of identifying shortcomings in current program to refine implementation - Product evaluation measures outcomes for decisions regarding continuation or refocus of programs - Key strategy is to work with carefully designed evaluation while maintaining flexibility - Evaluations should provide continuous stream of information to decision makers to make sure that programs continually improve
Tyler	Objectives Based Evaluation (1942)	Between Use & Methods	Questions and Methods	Program Oriented	- Major starting point for modern program evaluation - Primarily concerned with specification of objectives and measurement of outcomes - Steps include: 1. identify the purposes of education 2. select learning experiences that are useful for attaining objectives 3. organize these experiences 4. evaluate the effectiveness of the learning experiences

Weiss	Evaluation Research (1991)	Methods	Questions and Methods	NA	- Focus on traditional experimental methods - Recognizes evaluation as political activity - "decision accretion" – decisions are result of build-up of small choices, closing of options, and narrowing of alternatives

Wholey	Evaluation and Effective Public Management (1983)	Use	Questions and Methods	Decision Oriented	- Focus on managers and policymakers
					- Less concerned about stakeholders
					- Use of evaluation is to improve management
					- Four stage process for "sequential purchase of information"
					  (1) evaluability assessment – make initial assessment of extent to which it's feasible to conduct evaluation
					  (2) rapid feedback evaluation – focus primarily on extant data and easily collect information
					  (3) performance monitoring- measure program performance in comparison to prior or expected performance
					  (4) intensive evaluation – use comparison or control groups to gauge effectiveness of program activities in causing results

Wolf/Owens	Adversary Evaluation (Wolf, 1979) (Owens,	Valuing (2008)	Questions and Methods	NA	- no current advocates of this position

	1973)	Removed from current version			- suggest employing evaluators to represent two opposing viewpoints - These two then reach consensus on issues to be addressed by evaluation

Christie & Alkin Framework:

Methods – "grows from social inquiry foundational root. Primary focus is on developing models for evaluation practice which at the core are grounded in and derived from social science research methods. Models are mostly derivations of randomized control trial, intended to offer results that are generalizable and have a focus on 'knowledge construction'" (2012, p. 243)

Valuing – these theorists believe that the process of placing value on the evaluation is the essential component of an evaluator's work. "It is the work of the evaluator to make a judgment about the object that is being evaluated" (2012, p. 245). Branch also includes work of those interested in social justice in evaluation. Branch divided between post-positivists and constructivists

Use – These theorists are primarily concerned with the use of the evaluation itself and the information generated from the evaluation is used and focuses on those who will use the information

Stufflebeam (2012) Categorization:

Questions and Methods – Usually begin with a set of narrowly defined questions or predetermined method. Both approaches stress that it is more important to address a few pointed questions well than to try to do a broad assessment of merit or worth

Improvement and Accountability – The focus is on the assessment of a program's merit or worth, and are usually objectivist

Social Agenda and Advocacy - The focus is on ensuring that all segments of society have access to social and educational opportunities and services

Eclectic – Utilization-focused evaluation, which has elements of the other categories, but primarily focuses on ensuring that the evaluation has an impact

Fitzpatrick, Sanders, & Worthen (2011) Categorization:

Program-oriented – focus is on articulating goals and objectives and evaluating the extent to which they have been obtained

Decision-oriented – focus is on identifying and meeting needs of management decision makers

Consumer-oriented – goal is to develop evaluative information on "products" as well as accountability for use by consumers to help them choose among products and services

Expertise-oriented – depends on direct application of professional expertise to make judgment about quality of program being evaluated

Participant-oriented – involvement of stakeholders is central to determining the values, criteria, needs, and data necessary for the evaluation

APPENDIX C – APPROVAL TO CONDUCT 1:1 TECHNOLOGY PROGRAM EVALUATION

Catholic, College Preparatory
School for Young Women

January 22, 2012

Mrs. Anne Blake
20 Orchard Lane
Grosse Pointe Farms, MI 48236

Dear Mrs. Blake,

Thank you for the wonderful research proposal and presentation to the administrative team at ▮▮▮. The information was most valuable and the delivery was excellent. We have since met and I want to let you know that we are all willing to have ▮▮▮ High School participate in your research and that we believe the resulting information will help inform our technology plan moving forward.

We thank you for your sensitivity to the issue of instructional time and demands for time on those involved in the project. We look forward to working with you and wish you success with your very important research.

With highest regard,

Cheryl Delaney Kreger, Ed.D.
President

WOMEN WHO MAKE A DIFFERENCE

APPENDIX D – HUMAN INVESTIGATION COMMITTEE APPROVAL FORM

WAYNE STATE UNIVERSITY

IRB Administration Office
87 East Canfield, Second Floor
Detroit, Michigan 48201
Phone: (313) 577-1628
FAX: (313) 993-7122
http://irb.wayne.edu

NOTICE OF EXPEDITED APPROVAL

To: Anne Blake
Administration & Organization Stud

From: Dr. Scott Millis
Chairperson, Behavioral Institutional Review Board (B3)

Date: March 28, 2012

RE: IRB #: 036012B3E

Protocol Title: A Metaevaluation of the Impact Evaluation Process

Funding Source:

Protocol #: 1203010734

Expiration Date: March 27, 2013

Risk Level / Category: 45 CFR 46.404 - Research not involving greater than minimal riskResearch not involving greater than minimal risk

The above-referenced protocol and items listed below (if applicable) were **APPROVED** following *Expedited Review* Category (#7)* by the Chairperson/designee *for* the Wayne State University Institutional Review Board (B3) for the period of 03/28/2012 through 03/27/2013. This approval does not replace any departmental or other approvals that may be required.

- Revised Protocol Summary Form (received in the IRB Office 03/09/2012)
- Protocol (received in the IRB Office 03/09/2012)
- Receipt of letter of support from Mercy High School (dated 01/22/2012)
- The request for a waiver of the requirement for written documentation of informed consent has been granted for online survey according to 45 CFR 46.117(1)(2). Justification for this request has been provided by the PI in the Protocol Summary Form. The waiver satisfies the following criteria: (i) The only record linking the participant and the research would be the consent document, (ii) the principal risk would be potential harm resulting from a breach of confidentiality, (iii) each participant will be asked whether he or she wants documentation linking the participant with the research, and the participant's wishes will govern, (iv) the consent process is appropriate, (v) when used requested by the participants consent documentation will be appropriate, (vi) the research is not subject to FDA regulations, and (vii) an information sheet disclosing the required and appropriate additional elements of consent disclosure will be provided to participants not requesting documentation of consent.
- Internet Information Sheet
- Parental Consent Form (dated 03/01/2012)
- Written Assent for Minors (dated 03/01/2012)
- Data collection tools

- Federal regulations require that all research be reviewed at least annually. You *may* receive a *"Continuation Renewal Reminder"* approximately two months prior to the expiration date; however, it is the Principal Investigator's responsibility to obtain review and continued approval *before* the expiration date. Data collected during a period of lapsed approval is unapproved research and can never be reported or published as research data.
- All changes or amendments to the above-referenced protocol require review and approval by the IRB **BEFORE** implementation.
- Adverse Reactions/Unexpected Events (AR/UE) must be submitted on the appropriate form within the timeframe specified in the IRB Administration Office Policy (http://www.irb.wayne.edu//policies-human-research.php).

NOTE:

APPENDIX E – AGREEMENT TO CONDUCT EXPERT REVIEW OF METAEVALUATION

February 22, 2012

Anne,

I'd be happy to help with this project. I am out at meetings today but is there a good time to talk tomorrow am? Let me know a time and number where I can reach you. Mid morning is best for me.

Carl E. Hanssen

Office: 616-648-1290
Fax: 616-808-2866

1324 Lake Dr. SE
Grand Rapids, MI 49506

Dear Dr. Hanssen,
   My name is Anne Blake, and I am a doctoral candidate in Instructional Technology at Wayne State University. I am writing to you because I would like to hire you to conduct a metaevaluation as part of my dissertation research.
        The purpose of my research project is to study a new evaluation model proposed by Dr. Ingrid Guerra-Lopez (one of the professors in my program). In order to study the model, I am going to conduct an evaluation of a one-to-one laptop program at a secondary school here in Detroit using the model. Then, I am going to use Dr. Stufflebeam's Metaevaluation Checklist to evaluate the model. I have been in touch with Dr. Chris Coryn and have received permission to use the newly revised checklist. In order to triangulate my results, I am asking the stakeholders who participate in the laptop program evaluation to complete the Metaevaluation Checklist, too. Finally, I need to hire an expert evaluator to evaluate my metaevaluation...and that's where I hope you will come into the picture. I don't think it will take TOO much time, and I have a modest budget to compensate you.
        I hope you will consider my request thoughtfully. Please let me know if I can provide you with any additional information necessary to help you decide whether or not to take the job.
        Thanks and all the best!

APPENDIX F – INVITATION TO PARTICIPATE IN PROGRAM EVALUATION EMAIL

WXYZ High School High School Technology Program Evaluation
Request for Participation

Dear _____,

I am writing to ask you to participate in a research project that I am working on with WXYZ High School High School. The project has 2 parts:

(1) I am conducting a comprehensive evaluation of the laptop program at WXYZ High School. I will be using a new evaluation "model" (or process) to conduct the evaluation.

(2) I am conducting an "evaluation of the evaluation" as part of the research for my PhD.

I'm hoping that you will participate in this project because you represent one of the key stakeholder groups at WXYZ High School, the faculty. I know that your thoughtful contribution will strengthen the overall evaluation.

Here's what would be required of you if you agree to participate:

(1) Click on the link at the end of this email to indicate that you agree to participate

(2) Meet with me live or over the phone so that I can explain the process that I will be using to conduct the evaluation.

(3) Over the next 2 months, as I work through the process, I will either send you periodic emails, or call you, or meet with you in person (your choice) to update you on the evaluation and get your feedback and opinions on the laptop program and the evaluation process. I anticipate that we will talk approximately 3-4 times during this period.

(4) After I've collected the data for the evaluation, I will ask you to take an online survey to give me feedback about the process that I used to gather the information. I expect that I will send this to you in early June.

(5) When school re-opens in late August, I will send all of the participants a copy of my analysis of the laptop program, including recommendations for improvement based on the data that I gathered.

(6) Once you've read my report, I will ask you to take a second online survey to give me feedback about the report itself.

I anticipate that your participation in this project will take a total of 3-4 hours of your time over the next 4 months.

So, what's in it for you?
In addition to the satisfaction that you will get in knowing that you are supporting WXYZ High School High School, I will enter your name in a drawing for a new Kindle Fire. One participant in the evaluation project will win the Kindle.

The Fine Print:
This research is being conducted at Wayne State University. As a participant in this research study, there may be no direct benefit for you; however, information from this study may benefit the students at WXYZ High School High School, and elsewhere, in the future. There are no known risks at this time to participation in this study. There will be no costs to you for participation in this research study. You will be identified in the research records by a code name or number. Taking part in this study is voluntary. You are free to not answer any questions or withdraw at any time. Your decision will not change any present or future relationships with Wayne State University, WXYZ High School High School or their affiliates.

Questions:

If you have any questions about this study now or in the future, you may contact me at (313) 268-3908. If you have questions or concerns about your rights as a research participant, the Chair of the Human Investigation Committee can be contacted at (313) 577-1628. If you are unable to contact the research staff, or if you want to talk to someone other than the research staff, you may also call (313) 577-1628 to ask questions or voice concerns or complaints.

Please click HERE to verify your participation in the research study.

Sincerely,

Anne Blake, M.A., Ed. Spec., Doctoral Candidate
Wayne State University

APPENDIX G – INFORMED CONSENT FORM

Dear _____,

My name is Anne Blake and I am a doctoral student at Wayne State University. During the next few months I will be working with the faculty and staff of WXYZ High School High School to conduct an evaluation of the laptop program at the school. In addition, when the laptop program evaluation is complete, I will be conducting a "metaevaluation" (which is an evaluation of the evaluation) in order to complete the research for my doctoral dissertation.

I am writing to ask you to allow your daughter to participate in my research study. She has been selected because she has been recommended to me by a member of the faculty or administration. If you decide to allow your daughter to take part in the study, she will be asked to answer some informal questions about the laptop program, participate in email discussions during the course of the evaluation, and complete an online survey about the evaluation process at the conclusion of the project. The study will be conducted during April and May, 2012, and the final survey will be distributed in August, 2012. Your daughter has the option of not answering some questions, and she may withdraw from the study at any time. I anticipate that the entire project will take 3-4 hours of her time. You are welcome to view all of the materials that will be used during the study. You can contact me for copies of the evaluation materials at: ablake@wayne.edu.

There may be no direct benefits for your child; however, information from this study may benefit the students at WXYZ High School high school, and elsewhere, in the future. There are no known risks to your child for participation in this study. Needless to say, there are no costs to you or your child to participate in this study. For taking part in this research study, your daughter will receive a $20 iTunes gift card after she completes the online survey in August.

All information collected about your daughter during the course of this study will be kept confidential to the extent permitted by law. She will be identified in the research records by a code name or number. Information that identifies your child personally will not be released without your written permission. However, the Institutional Review Board (IRB) at Wayne State may review your child's records.

Your daughter's participation in this study is voluntary. You are free to withdraw her at any time. Your decision about allowing her to participate in the study will not change any present or future relationships with Wayne State University or its affiliates, WXYZ High School High School, your daughter's teachers, grades, or other services you or your child are entitled to receive.

If you have any questions about this study now or in the future, you may contact Anne Blake at (313) 268-3908. If you have questions or concerns about your rights as a research participant, the Chair of the Institutional Review Board can be contacted at (313) 577-1628. If you are unable to contact the research staff, or if you want to talk to someone other than the research staff, you may also call (313) 577-1628 to ask questions or voice concerns or complaints.

Consent to Participate in a Research Study:

To voluntarily agree to have your daughter take part in this study, you must sign on the line below. You are not giving up any of your or your child's legal rights by signing this form. Your

signature below indicates that you have read this entire consent form, including the risks and benefits, and have had all of your questions answered. You will be given a copy of this consent form.

_____

Name of Participant   Date of Birth

_____

Signature of Parent/ Legally Authorized Guardian   Date

_____

Printed Name of Parent Authorized Guardian   Time

_____

Signature of Person Obtaining Consent   Date

_____

Printed Name of Person Obtaining Consent   Time

APPENDIX H – INTRODUCTORY POWERPOINT PRESENTATION

Reach Consensus on Intended Results

- The evaluator will work with each group of stakeholders to identify the questions that the evaluation will attempt to answer
- All participants will have input about the evaluation questions

Third Step – Derive Measurable Indicators

- Once the evaluation questions have been identified, the evaluator will develop a tentative list of measurable indicators (observable phenomena) that seem to point towards the answers to the questions

- Example:
  - If a school wanted to know if their new attendance program was reducing absenteeism, a measurable indicator would be student attendance before and after program implementation

Measurable Indicators

- All participants will be asked to review the evaluation questions and the proposed measurable indicators and give the evaluator feedback about what is missing

Online Survey #1

- At this point in the project, you will be asked to answer a set of yes/no questions about the process used to conduct the evaluation

- Estimated time frame – June

- Now you can take a break while the evaluator collects and analyzes the data

Evaluation Results

- When the evaluator has completed the evaluation report, each participant will receive a copy and will be invited to give feedback/express concerns

- Estimated time frame – Late August/Early September

Online Survey #2

- After you have read the evaluation report, you will be asked to answer a few more yes/no questions online.

Overarching Values

- The evaluator is committed to providing the ███ Community with an ethical, useful, accurate, and comprehensive evaluation of the technology program

APPENDIX I – SURVEYS

Parent Survey

1. To what extent was the existence of a laptop/iPad program a factor in your decision to send your daughter to WXYZ High School High School?
____ It was the main reason I sent my daughter to WXYZ High School High School
____ It was a significant reason I sent my daughter to WXYZ High School
____ It was somewhat of a factor in my decision to send my daughter to WXYZ High School
____ It was a small factor in my decision to send my daughter to WXYZ High School
____ It did not play a role in my decision to send my daughter to WXYZ High School

2. The existence of an effective laptop/iPad program at WXYZ High School will influence my donation decisions in the future to things such as the annual fund, endowment fund, or capital campaigns.

____ Strongly Agree   ____ Neutral   ____ Disagree
____ Agree   ____ Strongly Disagree

3. The total cost of sending my daughter to WXYZ High School High School (including tuition and the cost of the laptop/iPad) is reasonable.

____ Strongly Agree   ____ Neutral   ____ Disagree
____ Agree   ____ Strongly Disagree

Recent Graduate Survey

1. What year did you graduate from WXYZ High School?
2009   2010   2011

2. The laptop program at WXYZ High School made school work more interesting.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

3. The laptop program at WXYZ High School improved the quality of my work.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

4. The laptop program at WXYZ High School improved my grades.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

5. I was more motivated to do schoolwork at WXYZ High School when I used my laptop.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

6. My laptop helped me to communicate with my teachers at WXYZ High School.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

7. The laptop program at WXYZ High School helped me learn how to analyze information.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

8. The laptop program at WXYZ High School helped me learn how to search for information.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

9. The laptop program at WXYZ High School helped me learn how to create presentations.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

10. The laptop program at WXYZ High School helped me learn how to work on assignments in small groups.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

11. The laptop program at WXYZ High School helped me learn how to organize information.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

12. The laptop program at WXYZ High School helped me learn how to work with a database.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

13. The laptop program at WXYZ High School helped me learn how to work with spreadsheets.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

14. The laptop program at WXYZ High School helped me learn how to work with social media.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

15. The laptop program at WXYZ High School helped me learn how to work with video editing and/or digital media.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

16. The laptop program helped me to develop critical thinking skills.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

17. The laptop program helped me to develop effective written communication skills.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

18. The laptop program helped me to develop effective oral communication skills.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

19. The laptop program increased my ability to access information.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

20. In terms of the ability to use technology, how did you compare to your peers when you began college?
_____ I was better prepared in the use of technology than my peers in college
_____ I was equally prepared in the use of technology as my peers in college
_____ I was less well-prepared in the use of technology as my peers in college
_____ I don't know

21. In terms of critical thinking skills (the ability to analyze information and draw conclusions about it) how did you compare to your peers when you began college?
_____ I have better critical thinking skills than my peers in college
_____ I have the same critical thinking skills as my peers in college
_____ I have poorer critical thinking skills than my peers in college
_____ I don't know

22. True/False: I was accepted at:
- My first choice of college
- My second choice of college

Faculty Survey

1. In which subject area do you teach?

English   Social Studies   Math   Language   Physical Education   Performing Arts/Music/Art
Religious Studies   Science   Other   Prefer Not to Say

2. Laptops help me to communicate with my students.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

3. How many "challenge-based learning" (CBL) projects did you participate in last year?

___0   ____ 1   ____ 2   ____ 3   ____ More than 3

4. If you participated in challenge-based learning projects (CBL) did they include a technology component, such as video or PowerPoint presentations?

____ Yes, all CBL projects included a technology component
____ Most included a technology component
____ Some included a technology component
____ None of the CBL projects included a technology component
____ I don't know

5. Overall, how often do you use each of the following?

	Never	Less than Monthly	Monthly	Weekly	Daily	I Don't Know
Moodle						
Online Quizzes						
Blogs						
Wikis						
Audacity						
Ning						
YouTube						
Google Docs						
Skype						
Facebook						
DyKnow						
PowerPoint						
Excel						
MovieMaker						

6. During the past school year I received sufficient professional support, which helped me to effectively integrate technology into my classes.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

7. With the change to iPads, I have received or am receiving sufficient professional development so that I am confident that I will be able to effectively integrate technology into my curriculum/classroom next year.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

8. When I have questions about integrating technology into my classes, I know where to go to get answers.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

9. The administration has reasonable expectations about the incorporation of technology into the curriculum.

Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

11. What effect has the laptop program had on student academic performance OVERALL?

	Declined	No Effect	Improved
Participation in Class			
Preparation for Class			
Attendance			
Behavior			
Motivation			
Engagement and interest level			
Ability to work in groups			
Ability to retain content material			
Quality of work			
Interaction with other students			

12. What effect has the laptop program had on student academic performance for AT-RISK or LOW ACHIEVING students?

	Declined	No Effect	Improved
Participation in Class			
Preparation for Class			
Attendance			
Behavior			
Motivation			
Engagement and interest level			
Ability to work in groups			
Ability to retain content material			
Quality of work			

	Declined	No Effect	Improved
Interaction with other students			

13. What effect has the laptop program had on student academic performance for HIGH ACHIEVING STUDENTS?

	Declined	No Effect	Improved
Participation in Class			
Preparation for Class			
Attendance			
Behavior			
Motivation			
Engagement and interest level			
Ability to work in groups			
Ability to retain content material			
Quality of work			
Interaction with other students			

14. If you have any comments that you would like to make about professional development (as it relates to the technology program) please provide them here.

Student Survey
1. Which year did you just complete?
_____ Senior   _____ Junior   _____ Sophomore   _____ Freshman

2. To what extent was the existence of a laptop program at WXYZ High School a factor in your decision to come to WXYZ High School?
_____ It was the main reason I came to WXYZ High School
_____ It was a significant reason why I came to WXYZ High School
_____ It was somewhat of a factor in my decision to attend WXYZ High School
_____ It was a small factor in my decision to attend WXYZ High School
_____ It did not play a role in my decision to attend WXYZ High School

3. Laptops make school work more interesting.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

4. Laptops make schoolwork easier to do.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

5. Laptops have improved the quality of my work.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

6. Having a laptop has improved my grades.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

7. I do more homework outside of school if I am able to use my laptop.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

8. I am more motivated to do schoolwork when I use my laptop.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

9. What I learn in school is relevant to my life right now.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

10. What I learn in school is helping me to prepare for the future.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

11. Laptops help me to communicate with my teachers.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

12. The laptop program has helped me learn how to analyze information.
Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

Last year, how often did you use your laptop to complete the following tasks?
13. Search for information.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

14. Create presentations and projects on my own.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

15. Work on assignments in small groups.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

16. Organize information.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

17. Work with a database.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

18. Work with a spreadsheet.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

19. Work with social media, such as Twitter, Facebook, Instagram
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

20. Take notes in class
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

21. Communicate using email or instant messages.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

22. Take a quiz, test, or assignment
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

23. Complete homework
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

24. Do drills to increase skills in Math, English, etc.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

25. Work on websites, digital, film/media, video editing, etc.
___Never   ___Less than Monthly   ___Monthly   ___Weekly   ___Daily

26. How many challenge-based learning (CBL) projects did you participate in last year?
___ 0   ___ 1   ___ 2   ___ 3   ___ More than 3

27. If you participated in challenge-based learning projects (CBL) did they include a technology component, such as video or PowerPoint presentations?
____ Yes, all CBL projects included a technology component
____ Most included a technology component
____ Some included a technology component
____ None of the CBL projects included a technology component
____ I don't know

28. Overall, how often did your teachers use the following:

	Never	Less than Monthly	Monthly	Weekly	Daily	I Don't Know
Moodle						
Online Quizzes						
Blogs						
Wikis						
Audacity						
Ning						
You Tube						
Google Docs						
Skype						
Facebook						
DyKnow						
PowerPoint						
Excel						
MovieMaker						

APPENDIX J – FINAL EVALUATION REPORT ON TECHNOLOGY PROGRAM

WXYZ High School Student Computing Program Evaluation Fall, 2012

Executive Summary

In 2006, after a one year pilot program, WXYZ High School introduced a one-to-one student computing program. The program was, in part, a response to the board of directors' call to move the school from viability to vitality. As part of the program, students were required to purchase a tablet computer. For the current school year (2012-2013) the school has begun a transition to iPads. This evaluation was conducted to determine if the technology program has helped WXYZ to achieve the results that they sought when the program was implemented, and to strengthen the program during the transition to iPads. A group of more than 30 representative stakeholders participated in the planning and execution of the evaluation, which is being done as part of a doctoral research project at Wayne State University.

The results of the evaluation reveal many positive trends for WXYZ. Some of these trends are directly attributable to the technology program, and some are not. The good news includes the following: (a) the school has seen an increase in its Composite ACT scores during the past ten years as well as increases in several ACT subject scores, including Math, English, and Science, (b) the school has seen an increase in the PSAT critical reading score, (c) on a percentage basis, WXYZ lost fewer students than its competitors during the recent recession and is the only school among its competitors to see an increase in enrollment in the past several years, (d) the school has seen an increase in market share in the past 5 years, (e) the school has seen an increase in the number of college scholarships received per student during the past nine years, (f) the school's overall donation levels remained strong throughout the recent recession, (g) students have attended increasingly competitive colleges over the past ten years, (h) the school has seen an increase in the number of students pursuing STEM-related fields of study. Additional positive results are explained in detail in the body of the report. The evaluator recommends that the school use this data as a marketing tool in the future.

There were a few areas of concern that came to light as a result of the evaluation. These include the following: (a) the school has seen a decrease on the PSAT Writing Skills scores over the past ten years, (b) while the faculty is supportive of the technology program, they have not fully integrated technology into their curriculum, (c) the faculty believes that the program has had a negative effect on student participation and behavior, and (d) low achieving and at-risk students, in particular, have not experienced as many positive effects as a result of the program.

Based on these areas of concern, the evaluator recommends that the school focus on providing additional, targeted support for faculty efforts at integrating technology. Teachers should be required to show evidence of technology integration, but should also be supported by a specialist who works with them to find effective ways to use it in individual classrooms. A technology integration specialist should also be able to support faculty efforts towards improving the achievement level of at-risk students. The faculty and administration must work together to directly address the problem of student behavior and participation. In spite of these concerns, the student computing program appears to have contributed significantly to WXYZ's progress towards vitality.

1. BACKGROUND FOR THE EVALUATION

1.1 Introduction

In September 2006, after a one year pilot program, WXYZ in Farmington Hills, Michigan introduced a one-to-one computing program which included the requirement that all incoming freshmen purchase an HP tablet computer. The program was designed to facilitate the integration of technology into the curriculum and school operations, as directed by the school's board of directors.

This evaluation of the one-to-one computing program was conducted by a doctoral student in Instructional Technology at Wayne State University as part of a larger research project. The evaluation was conducted at the end of the sixth full year of the program, and coincides with the school's decision to convert from the HP tablet computer to the iPad. This report outlines the findings of the evaluation.

1.2 History and Status of the Program

The WXYZ Student Computing Program developed as a response to a directive from the school's board of directors beginning in 1998 to move WXYZ from "viability to vitality". The 1998 Technology Vision and Goals statement included the following objectives: (1) enhance the teaching/learning process with a variety of technologies, (2) enhance the learning environment to support teachers, facilitate a shift towards learning as the construction of knowledge, promote the concept of lifelong learning, and promote the integration of values into teaching/learning, (3) integrate technology throughout the curriculum, (4) extend the boundaries of the learning environment beyond the school building, and (6) assess the impact of technology on learning. During the 2005-2006 school year the school conducted a pilot program, in which 50-60 students used laptops throughout the year. Beginning with the 2006-2007 school year, all incoming freshman have been required to purchase laptops, at a cost of approximately $2,200 per student.

In 2011, an internal evaluation of the program was conducted, which included a staff usage survey, parent and student focus groups, and meetings with IT specialists. The decision was made to change from the tablet computer to iPads for incoming students beginning with the 2012-2013 school year. In late 2011 the evaluator approached the school's administration to request that the school participate in an evaluation of the computing program as part of another research project. The school agreed to participate and provided the evaluator with full access to stakeholders and resources.

1.3 Evaluation Overview

1.3.1 Purpose

The evaluation has two primary purposes:

1. To determine if the technology program is helping WXYZ to achieve the results that it sought when the program was implemented.

2. To help WXYZ identify ways to strengthen the program as it transitions to iPads.

1.3.2 Desired Results of Program

The representative stakeholders who participated in the evaluation identified twelve desired results from the technology program. These results are divided into three levels and are listed in Table 1.

Table 1. Results Sought From Student Computing Program

Level	Results Sought
1. Internal	1.1 The Student Computing Program will help students to develop college and 21st Century skills, particularly: - critical thinking skills - effective written and oral communication skills - the ability to access information - the ability to analyze information - problem solving skills - organizational skills
	1.2 The Student Computing Program will help to increase competence and confidence among students in the creation and use of information, particularly in the use of technology for: - word processing - creating presentations - database management - spreadsheet management - video production and editing - social media - the use of "apps"
	1.3 The Student Computing Program will help to enrich the classroom experience and increase motivation by making learning more enjoyable and by engaging students in knowledge creation and use
	1.4 The Student Computing Program will allow the faculty to implement curriculum more effectively and will facilitate the individualization of instruction
	1.5 The Student Computing Program will facilitate communication between students and faculty
2. Organizational	2.1 The Student Computing Program will help W to distinguish itself from other schools in the area and facilitate marketing
	2.2 The Student Computing Program will help WXYZ move "from viable to vital" (Board directive)
	2.3 The Student Computing Program will incorporate the global shift in emphasis from learning as an "assimilation of facts" to learning as "knowledge creation and use by the individual"

	2.4 The Student Computing Program will help to extend the boundaries of the learning environment beyond the school building and 8 a.m. – 3 p.m.
	2.5 The Student Computing Program will promote the integration of values into teaching and learning (from the Technology Vision Statement)
3. Societal	3.1 The Student Computing Program will help WXYZ to educate women who make a difference (from the School Motto)
	3.2 The Student Computing Program will help to foster the spiritual, intellectual, moral, physical, and cultural development of the students (from the Mission Statement)

1.3.3 Evaluation Questions

The evaluation questions stem from the desired results that were identified by the stakeholders of the program, and are listed below:

1. Has the program helped students to develop college and 21st Century skills?

2. Has the program increased students' competence and confidence in the creation and use of information?

3. Has the program enriched the classroom experience and increased motivation by making learning more enjoyable and engaging students in their own knowledge creation and use?

4. Has the program helped the faculty to implement the curriculum more effectively and facilitated the individualization of instruction?

5. Has the program facilitated communication between students and faculty?

6. Has the program helped WXYZ distinguish itself from other schools in the area and has it facilitated marketing?

7. Has the program helped WXYZ move from "viable to vital"?

8. Has the program incorporated the global shift in learning towards the concept of "knowledge creation and use by the individual"?

9. Has the program extended the boundaries of the learning environment beyond the school building and 8 a.m. – 3 p.m.?

10. Has the program promoted the integration of values into teaching and learning?

11. Has the program helped WXYZ to educate women who make a difference and foster the spiritual, intellectual, moral, physical, and cultural development of its students?

2. Methods

The evaluation was conducted using a framework known as the Impact Evaluation Process that includes the following steps. First, all stakeholder groups were identified. A stakeholder is defined as anyone who can affect the Student Computing Program or can be affected by it. Representatives of each of these groups were invited to participate in the program evaluation. The stakeholder groups include: the administration, the faculty, students, parents, the board of directors, the Sisters of WXYZ, recent graduates of WXYZ, prospective WXYZ students, and members of the Information Technology staff. In all, more than 30 individuals participated in the planning of the evaluation.

Next, the participating stakeholders individually identified the results that they expect to see as a result of the Student Computing Program. The representatives of the stakeholder groups reached consensus about the expected results at each of three levels: internal, organizational, and societal. Internal results are those that affect the students directly. Organizational results affect WXYZ as a whole. Societal results reflect WXYZ's potential contribution to society.

Finally, the participating stakeholders identified and reached consensus about measurable indicators that point towards the desired results. The evaluator identified data sources for each of the measurable indicators. The evaluator collected the data from the identified data sources and analyzed it for this report.

2.1 Specific Results and Indicators to Observe

In order for an evaluation to provide valid and meaningful information, it must be based on relevant, reliable information. For each of the evaluation questions, the participating representative stakeholders reached consensus on measurable indicators that point towards the desired results. These indicators are shown below in Table 2.

Table 2. Measurable Indicators for Each Evaluation Question

Evaluation Question	Measurable Indicators
1. Has the program helped students to develop college and 21st century skills?	• Changes in standardized test scores • Current students' perceptions • Recent graduates' perceptions
2. Has the program increased students' competence and confidence in the creation and use of information?	• Current students' perceptions • Recent graduates' perceptions • Implementation of challenge-based learning
3. Has the program enriched the classroom experience and increased motivation by making learning more enjoyable and engaging students in their own knowledge creation and use?	• Current students' perceptions • Recent graduates' perceptions • Faculty perceptions • Implementation of challenge-based learning
4. Has the program helped the faculty to	• Changes in standardized test scores

implement the curriculum more effectively and facilitated the individualization of instruction?	• Faculty professional development • Slice impact – lower and higher students should benefit most • Faculty perceptions
5. Has the program facilitated communication between students and faculty?	• Current students' perceptions • Faculty perceptions • Use of technology based communication tools
6. Has the program helped WXYZ distinguish itself from other schools in the area and has it facilitated marketing?	• Computing program as a factor in attracting new students • Computing program at WXYZ compared to similar schools
7. Has the program helped WXYZ move from "viable to vital"?	• Enrollment trends • Market share compared to similar schools • Faculty attraction/retention • Donation/Endowment trends • Maintain competitive/affordable tuition • Changes in standardized test scores
8. Has the program incorporated the global shift in learning towards the concept of "knowledge creation and use by the individual"?	• Implementation of challenge based learning • Faculty development and use of unique technology-based teaching methods • Level of technology integration in all areas of curriculum
9. Has the program extended the boundaries of the learning environment beyond the school building and 8 a.m. – 3 p.m.?	• Faculty use of unique technology-based teaching methods • Level of technology integration in all areas of curriculum
10. Has the program promoted the integration of values into teaching and learning?	• Movement towards green technologies such as textbook-less classrooms • Development and application of technology acceptable use policies and cyberbullying policies
11. Has the program helped WXYZ to educate women who make a difference and foster the spiritual, intellectual, moral, physical, and cultural development of its students?	• Number of scholarships awarded to graduates • College attendance rates • Quality of colleges attended • Number of students who pursue majors/careers in Science, Technology, Engineering, or Math (STEM fields)

2.2 Data Sources and Instruments/Methods of Data Collection

2.2.1 Current Student Survey

The student survey addresses questions about reactions, knowledge, attitudes, and skills. The survey consists of 30 items, most of which have a 5-point, Likert rating scale. The scale ranges from "strongly agree" to "strongly disagree". Responses were evenly split among all grade levels. With 209 respondents from a population of 750 (27.9% response rate), the survey results compute to 5.8% margin of error at 95% confidence level. Therefore, sample results are representative of the population.

2.2.2 Recent Graduate Survey

The recent graduate survey addresses questions about reactions, knowledge, attitudes, skills, and college preparedness. The survey consists of 22 items, most of which have a 5-point, Likert rating scale. The scale ranges from "strongly agree" to "strongly disagree". Responses were collected from the two classes of graduates who fully participated in the laptop program at WXYZ. 37% of respondents were from the class of 2010. 63% of respondents were from the class of 2011. Although the population of these two classes is 376, the survey was only sent to 221 students because of availability of email addresses. With 69 responses from a population of 376 (18.4% response rate), the survey results compute to a 10.5% margin of error at 95% confidence level. This high margin of error needs to be considered when evaluating the results.

2.2.3 Parent Survey

The parent survey addresses questions about attitudes and reactions. The survey consists of three items. With 153 responses from a population of approximately 700 (21.8% response rate), the survey results compute to a 7% margin of error at 95% confidence level. Therefore, sample results are representative of the population.

2.2.4 Faculty Survey (2012)

The faculty survey addresses questions about reactions, attitudes, and skills. The survey consists of 13 items. Responses were split among departments as is depicted in Figure 1. With 34 responses from a population of 45 classroom teachers (75.6% response rate), the survey results compute to a 8.5% margin of error at 95% confidence level. This margin of error is at the high end of the acceptable range and needs to be considered when evaluating the results.

Figure 1 - Faculty Survey Participation by Department
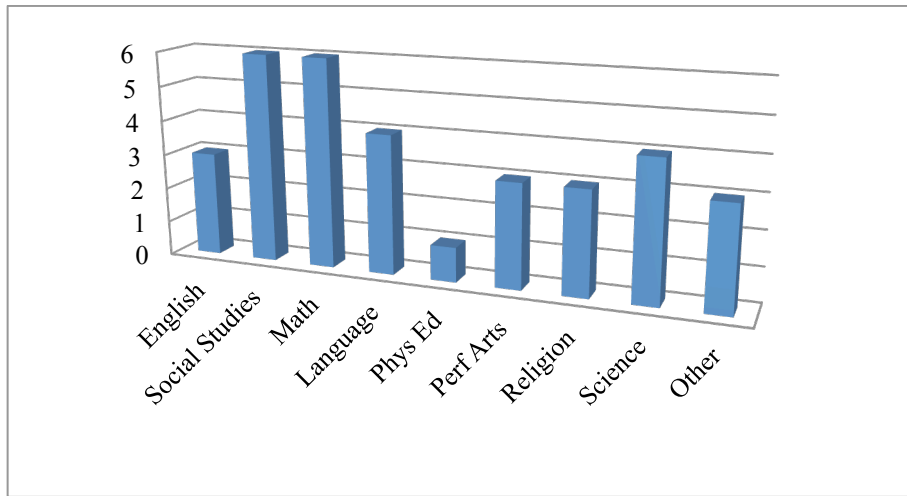
2.2.5 Other Data Sources
Other data sources used in the evaluation include:
- Survey of faculty conducted in 2011
- ACT scores for past ten years
- PSAT scores for past ten years
- College scholarship data for past ten years
- College choice data for past ten years
- Technology department professional development schedule
- Donation/endowment data for past ten years
- Student Agenda Book
- Enrollment data for school and archdiocese for past ten years
- Extant data about competitor schools found on school websites

2.3 Data Collection Schedule
The schedule of major data collection activities is summarized in Table 3.

Table 3. Data Collection Schedule

Data Collection Activity	Dates
Representative Stakeholders Invited to Participate in Evaluation	April, 2012
Consensus Reached on Expected Program Results and Measurable Indicators	June, 2012
Administration of student survey	June, 2012
Administration of recent graduate survey	July, 2012
Administration of parent survey	July, 2012
Administration of faculty survey	July, 2012

2.4 Discussion of Relevant Factors

Interviews with representative stakeholders were conducted primarily via in person meetings, email, and phone conversations. Results of phone interviews were transcribed and results categorized by theme. The responses to open-ended survey questions were analyzed in a similar manner. The rating scale survey items were analyzed and reported using descriptive statistics.

The potential for bias in the results of this evaluation exists due to the fact that the evaluator has three children who have attended or currently attend the school. This potential for bias is addressed by hiring a professional evaluator to review the results of the evaluation after initial review by all stakeholders.

WXYZ incurred no costs associated with the evaluation. The evaluator received partial funding from the university with which she is associated. The remainder of the cost was borne by the evaluator as part of a larger research project.

3. Findings and Interpretations

3.1 College Readiness and Twenty-First Century Skills

The student computing program should help students develop college skills and 21st century skills. In particular, as a result of the program there should be an improvement in critical thinking skills, oral and written communication skills, the ability to access information, the ability to analyze information, and organizational skills.

3.1.1 ACT Scores as Indicator of College Readiness

The ACT test measures high school students' general educational development and capability to complete college level work. According to the ACT organization, scores provide an indicator of college readiness. An increase in ACT scores over time is, therefore, an indicator of improvement in college readiness. Figure 2 illustrates the Composite ACT scores for WXYZ students over the past ten years. The graph also shows the Composite ACT scores for students in the state of Michigan. The vertical line indicates the year that the first students who participated in the student computing program took the ACT test. While there has been a trend toward higher ACT scores over the past 10 years, and particularly in the past two years, it is impossible to conclude that the computing program is the cause of the improvement.

Figure 2 - Composite ACT Scores
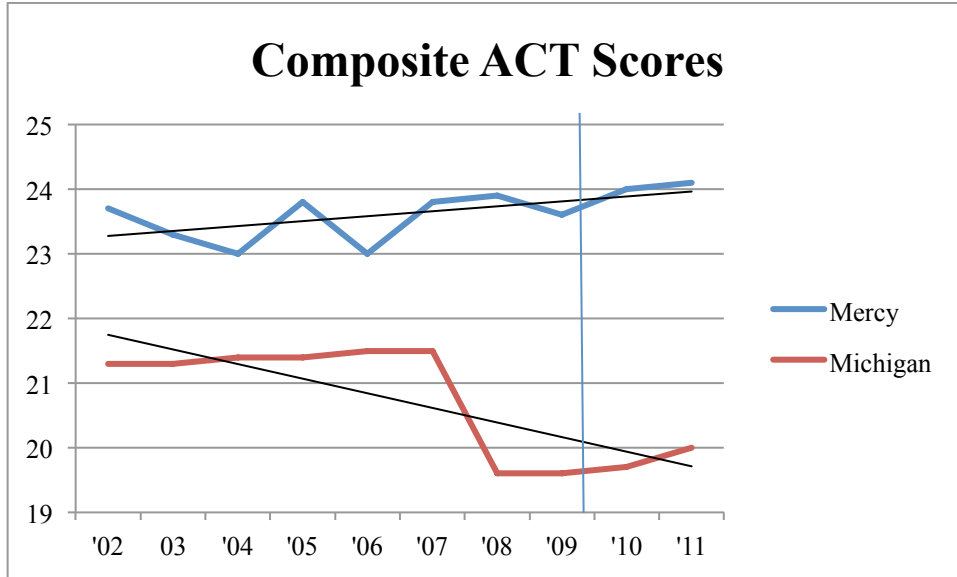
WXYZ has also seen a trend towards higher ACT scores in most of the subject tests (Math, English, Science) that make up the composite score, as seen in Figures 3-5. The only subject that has not improved over time is Reading, as shown in Figure 6.
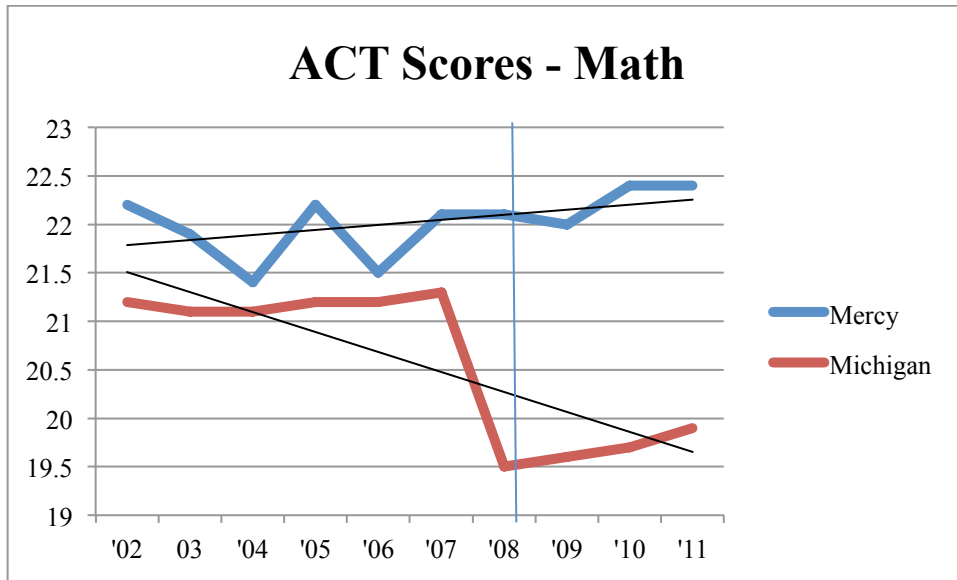
Figure 3 - ACT Scores – Math
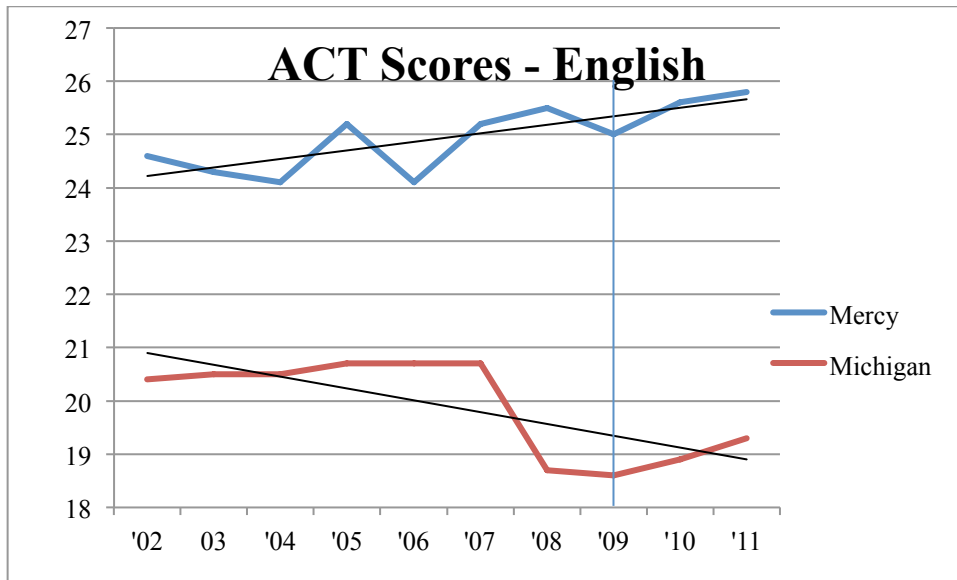
Figure 4 - ACT Scores – English
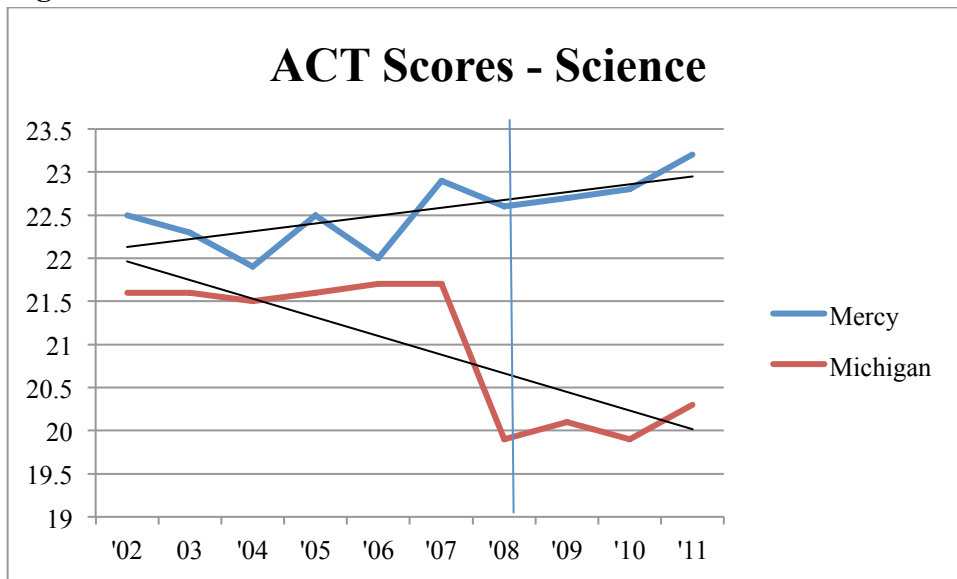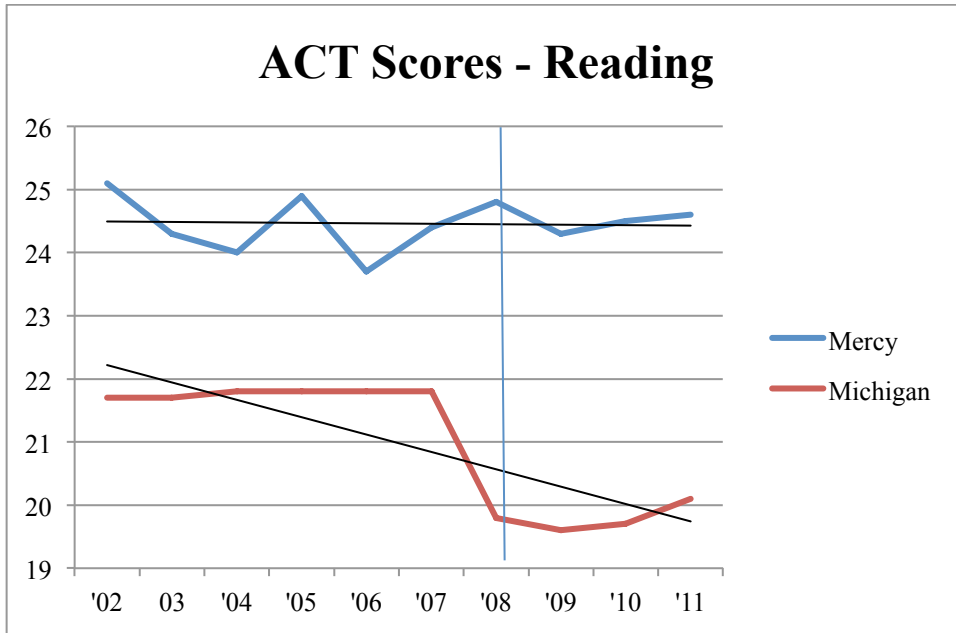

Figure 5 - ACT Scores – Science

Figure 6 - ACT Scores – Reading


3.1.2 Written and Oral Communication Skills

Skills in written and oral communication are indicators of college readiness. The PSAT includes an assessment of writing skills, and is one of the few negative trends at WXYZ. Figure 7 shows the average PSAT writing skills scores for the past ten years. Although there is no causal link between the technology program and the decrease in writing skills scores, it is an area in which there is the potential for growth in the future.

Figure 7 - PSAT Writing Skills Scores
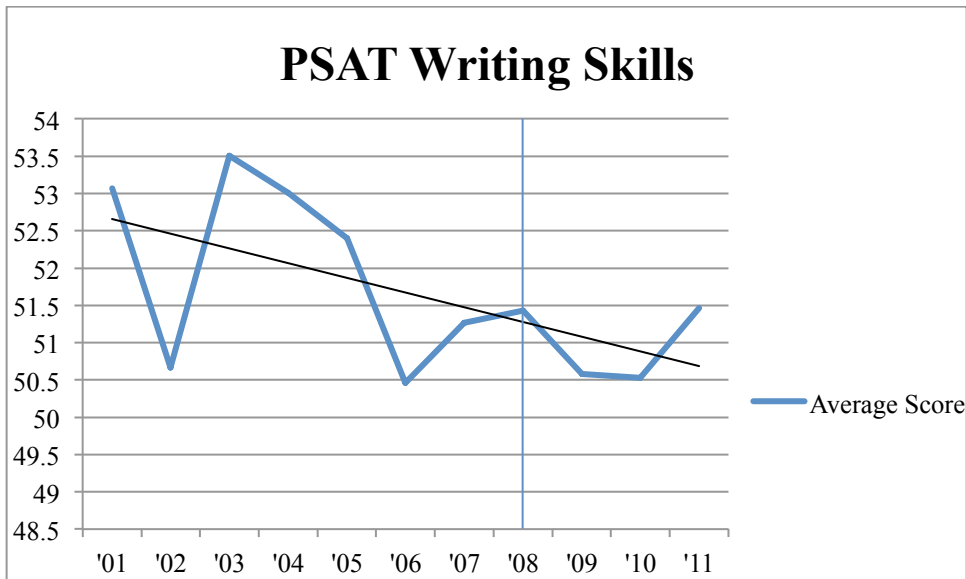
In spite of this negative trend, 65% of recent graduates strongly agree or agree that the laptop program helped them to develop effective written communication skills, as shown in Figure 8. However, the majority of recent graduates did not feel that the program affected their oral communication skills.

Figure 8- Recent Graduates' Perception of Program Effect on Communication Skills
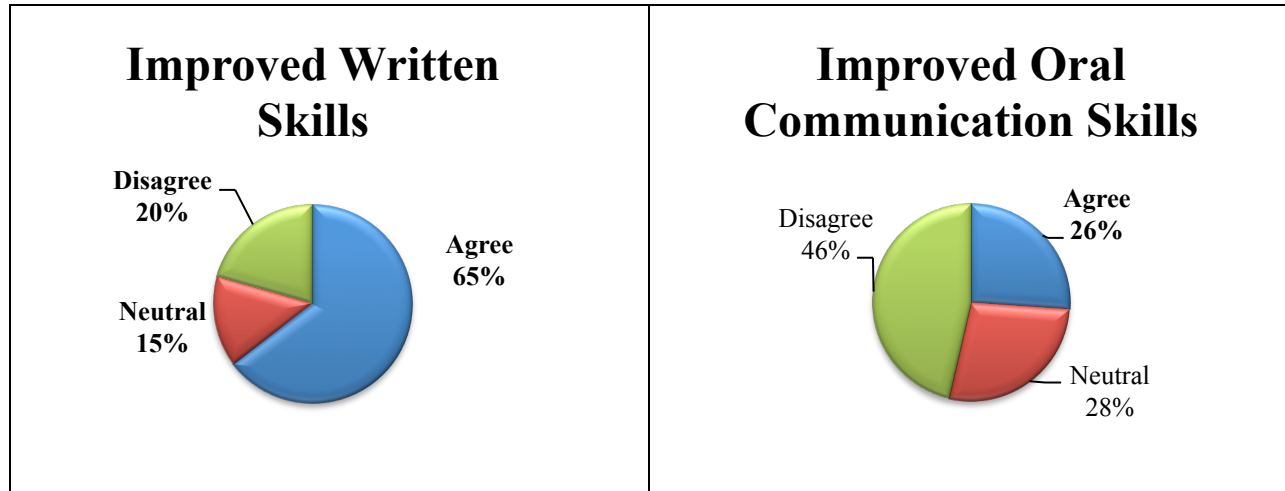
3.1.3 Ability to Access Information

As information is increasingly available through the use of technology, the ability to access it has become an indicator of preparedness for life in the 21st Century. 97.5% of current students indicated that they use their laptops daily or weekly to search for information. An overwhelming number of recent graduates indicated that the laptop program increased their ability to access information, as shown in Figure 9.

**Figure 9 –Program Effect on Recent Graduates' Ability to Access Information**
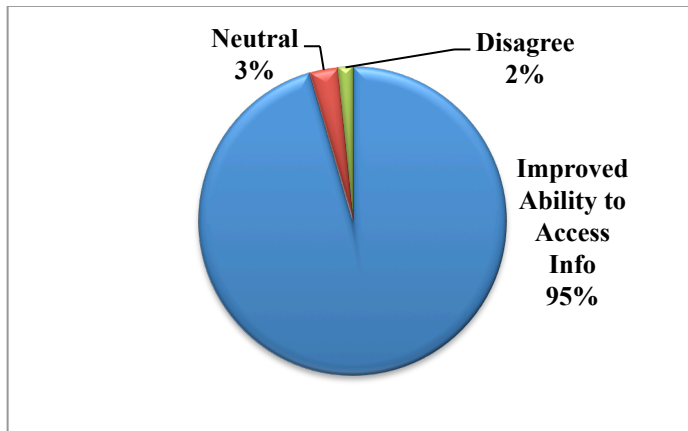
3.1.4 Ability to Analyze Information

After accessing information, it is important that students learn how to analyze it. The ability to analyze information is a critical 21st Century skill. This skill will be addressed again in the upcoming section of this report that deals with challenge-based learning. While it is difficult to

quantitatively evaluate the ability to analyze information, it is worthwhile to assess the students' perceptions of their ability in this regard. 62% of current students believe that the laptop program has helped them learn how to analyze information. 42% of recent graduates believe that the laptop program helped them learn how to analyze information, as seen in Figure 10.

Figure 10 – Program Effect on Ability to Analyze Information
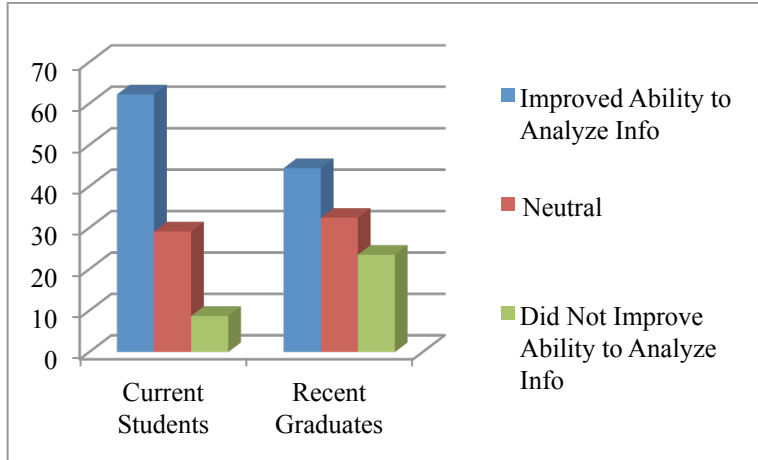
3.1.5 Organizational Skills

The ability to organize information is a critical skill for success in the 21st Century. 86% of current students indicate that they use their laptops daily or weekly to organize information. 68% of recent graduates indicate that the laptop program helped them learn how to organize information, as seen in Figure 11.

Figure 11 – Program Effect on Recent Graduates' Ability to Organize Information
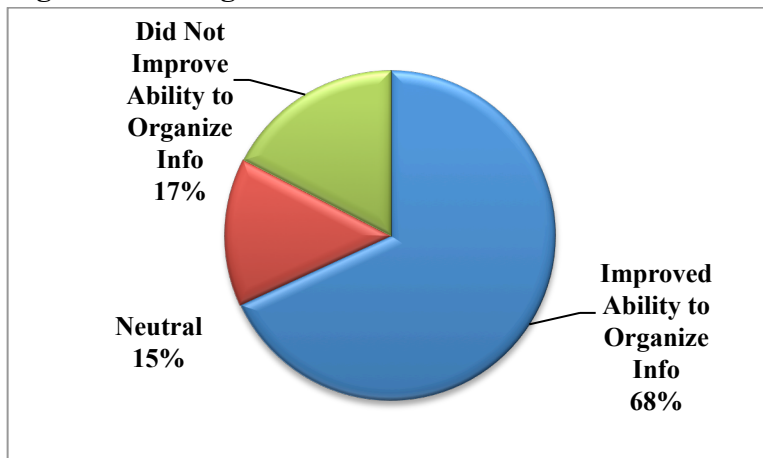
3.1.6 Critical Thinking Skills

Critical thinking skills are notoriously difficult to assess quantitatively. The PSAT test includes a measure of critical reading skills, which includes measures of the meaning of words, reasoning and inference, organization, and ideas. WXYZ has seen an increase in PSAT critical reading scores over the past ten years, as shown in Figure 12. The vertical line indicates the first year that the students who participated in the computing program took the PSAT. While there has been a trend toward higher PSAT scores over the past 10 years, and particularly in the past two years, one cannot conclude that the computing program is the cause of the improvement.

Figure 12 - PSAT Critical Reading Skills
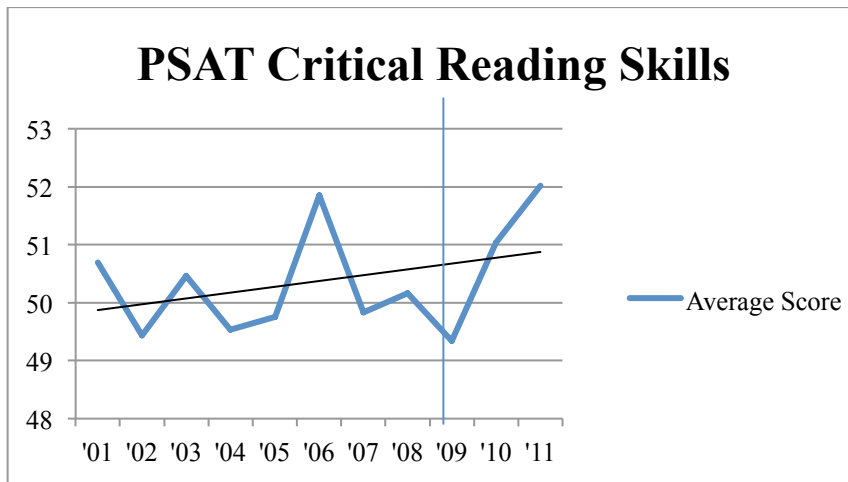
Current student and recent graduate perceptions are less reliable indicators of changes in critical thinking skills. While recent graduates gave mixed responses about the effect that the laptop program had on their critical thinking skills, they felt strongly that they were better prepared, or as well prepared, as their peers when they began college. Recent graduate perceptions are illustrated in Figures 13 and 14.

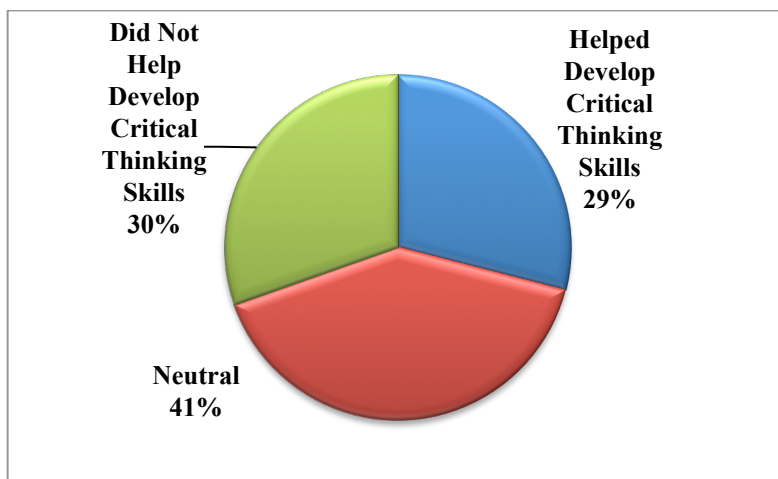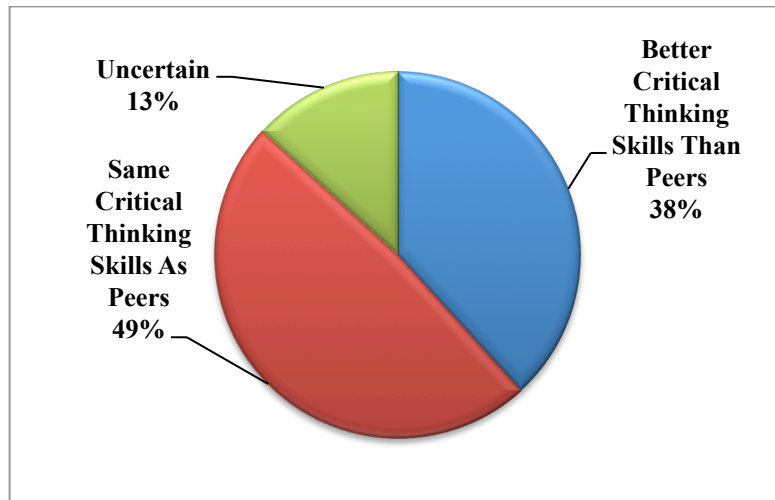**Figure 13 – Perceived Program Effect on Recent Graduates' Critical Thinking Skills**

Figure 14 - Recent Graduate Perceptions of Critical Thinking Skills vs. College Peers


3.2 Confidence and Competence in Creation and Use of Information

The student computing program should increase confidence and competence among students in the creation and use of information, particularly in the use of technology for creating presentations, video production and editing, database management, spreadsheet management, word processing, and the use of social media.

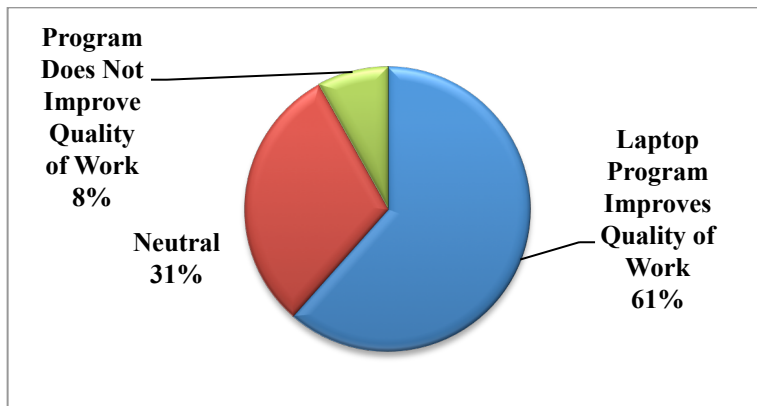### 3.2.1 Confidence in Creation and Use of Information

The majority (53%) of recent graduates strongly agreed or agreed that the laptop program improved the quality of their work. In addition, almost all recent graduates felt that they were better prepared or equally prepared in the use of technology as their peers at college, as seen in Figure 15. No recent graduates felt less well-prepared in the use of technology than their peers in college.

Figure 15 – Recent Graduates' Ability to Use Technology Compared to Peers in College
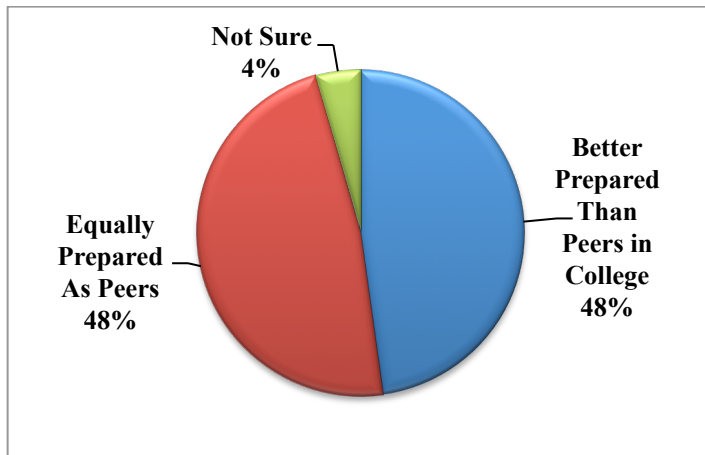
61% of current students strongly agreed or agreed that the laptop program has improved the quality of their work, as shown in Figure 16.

Figure 16 – Current Students' Perception of Program Effect on Quality of Work

3.2.2 Competence in Creation and Use of Information

Recent graduates felt strongly that the computing program contributed to their competence with a variety of resources critical to the creation and use of information. Figure 17 shows recent graduates' perception of the effect of the laptop program on their ability to create presentations, work with databases, create spreadsheets, work with social media, and video/digital editing. The strongest effects are seen in the ability to create presentations, work with databases, and work with social media. There is room for improvement in creating spreadsheets and working with video/digital editing.

Figure 17 – Recent Graduates' Perception of Program Effect on Competence in Creation & Use of Information
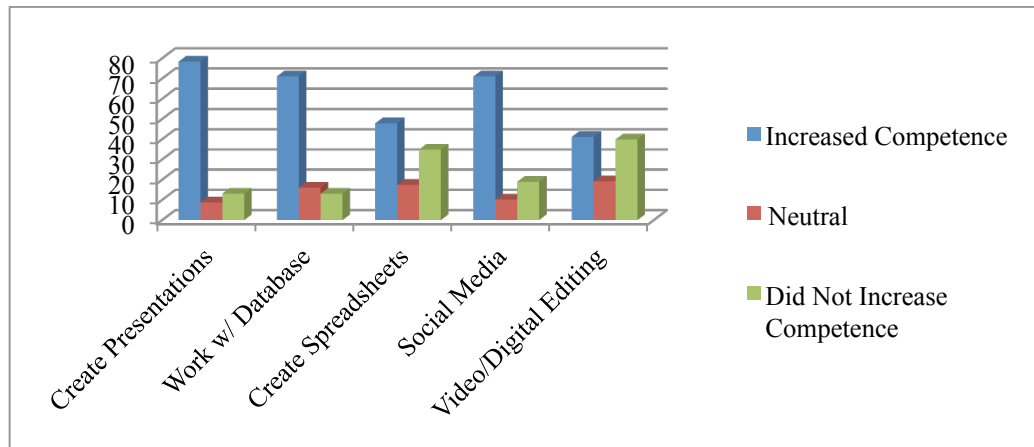
For current students, it is intuitive that frequent opportunities to work with these resources will ultimately impact their level of competence. While students indicate overwhelming daily use of social media, and fairly regular creation of presentations and opportunities to work with databases, it is apparent that opportunities to use spreadsheets and video/digital editing are less frequent as indicated in Figure 18.

Figure 18 – Current Students' Frequency of Use of Resources for Creation and Use of Information
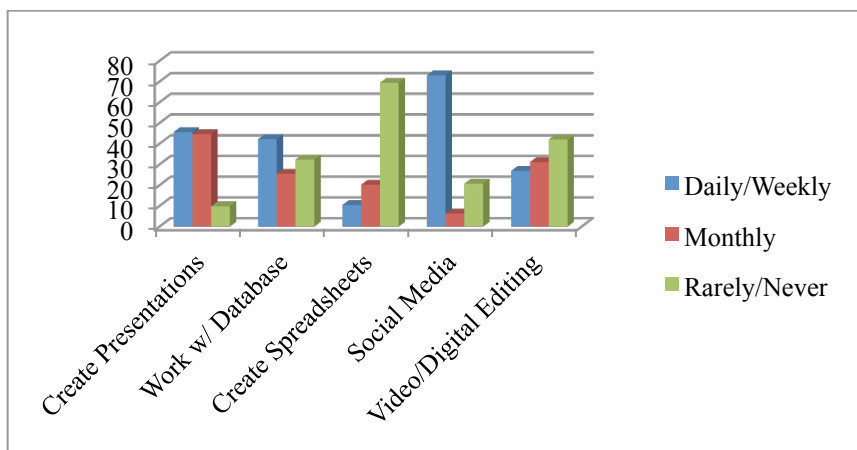
3.3 Enriched Classroom Experience and Increased Motivation and Engagement

The student computing program should enrich the classroom experience and increase motivation by making learning more enjoyable and engaging students in knowledge creation and use. The faculty expressed mixed views about the effect of the student computing program on the classroom experience. The majority of faculty members feel that the program has improved the students' level of preparedness for class. However, they believe that there has been a

corresponding decline in participation and behavior as a result of the program as illustrated in Figure 19. These results are a key area of concern.

Figure 19 – Faculty Perception of Laptop Effect on Classroom Engagement & Motivation
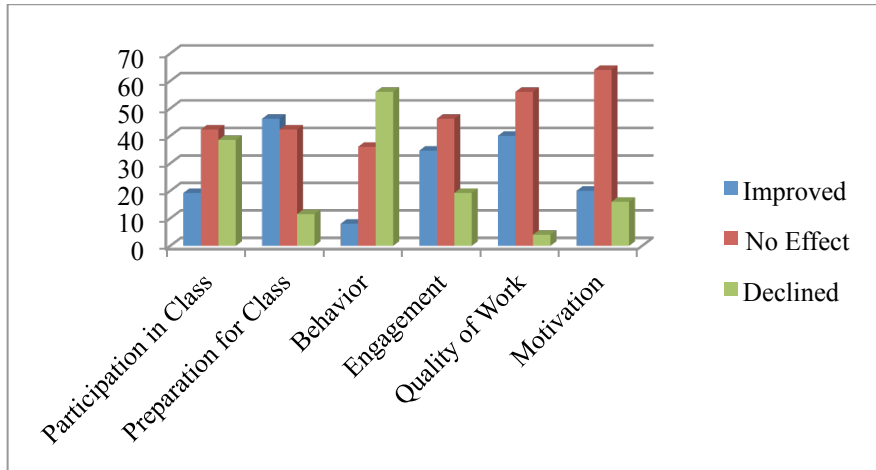
Current students overwhelmingly believe that laptops make school work easier to do and more interesting. They also believe that the laptop program has caused an improvement in the quality of their work. Ironically, however, they do not believe that they are more motivated to do schoolwork when they use their laptop. These results are illustrated in Figure 20.

Figure 20 – Current Student Perception of Laptop Effect on Schoolwork
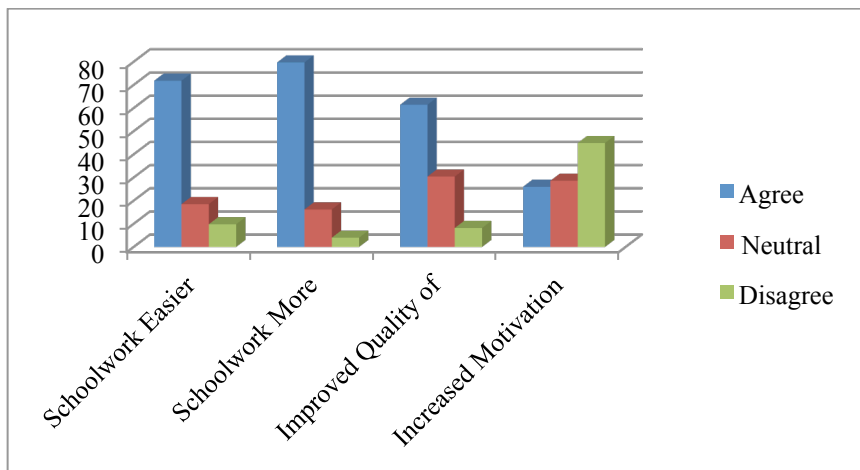
3.4 Faculty Ability to Effectively Implement Curriculum and Facilitate Individualized Instruction

3.4.1 Ability to Effectively Implement Curriculum

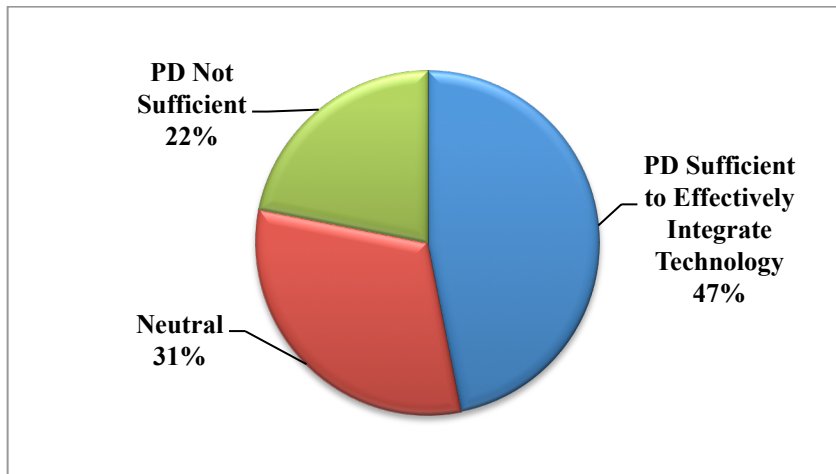The student computing program should help the faculty to implement the curriculum more effectively and should facilitate their ability to individualize instruction. Improvements in standardized test scores over time are an indicator that the faculty is successfully implementing the curriculum. However, as stated earlier, there is no proven causal link between the improvement in scores and the student computing program. An additional indicator of the faculty's ability to effectively implement the curriculum is the amount of professional development they receive that is targeted toward the integration of technology. One of the criticisms expressed by the faculty about the introduction of the laptop program seven years ago was that they did not receive sufficient training prior to program implementation. The administration has taken a different approach as it introduces the iPad program for Fall 2012. Throughout the spring and summer, faculty members have been offered a variety of workshops and labs that are designed to familiarize them with the iPad and Google-based tools. While some faculty members continue to express concern about not having enough time to learn how to use the iPads effectively, the majority of faculty members feel that they have received or are receiving sufficient professional development to effectively integrate technology into their curriculum, as seen in Figure 21. In addition, an overwhelming majority of faculty members (75%) indicated that they know where to go to get answers when they have questions about integrating technology into their classes.

When survey results are cross-tabbed by subject matter, more interesting information emerges. A relatively high percentage of faculty members in the English, Language, and Performing Arts departments felt that they did not receive sufficient professional support in order to effectively integrate technology into their curriculum during the past year. Correspondingly, these departments and the Math department expressed concern about receiving sufficient professional support during the conversion to iPads.

Figure 21 – Overall Faculty Perception of Professional Development

3.4.2 Facilitate Individualized Instruction

The student computing program should facilitate the faculty's ability to individualize instruction to meet the needs of different types of learners. In particular, the program should support students at both ends of the achievement spectrum, including high achievers and at-risk students. The faculty perceives that the program has improved high achievers' participation, level of preparation, engagement and quality of work. However, they believe that the program has caused a decline in participation and behavior for at-risk students. Therefore, it appears that the program has facilitated instruction for high achievers, but that there is room for improvement in its support of at-risk students.

Figure 22. Facilitation of Individualized Instruction
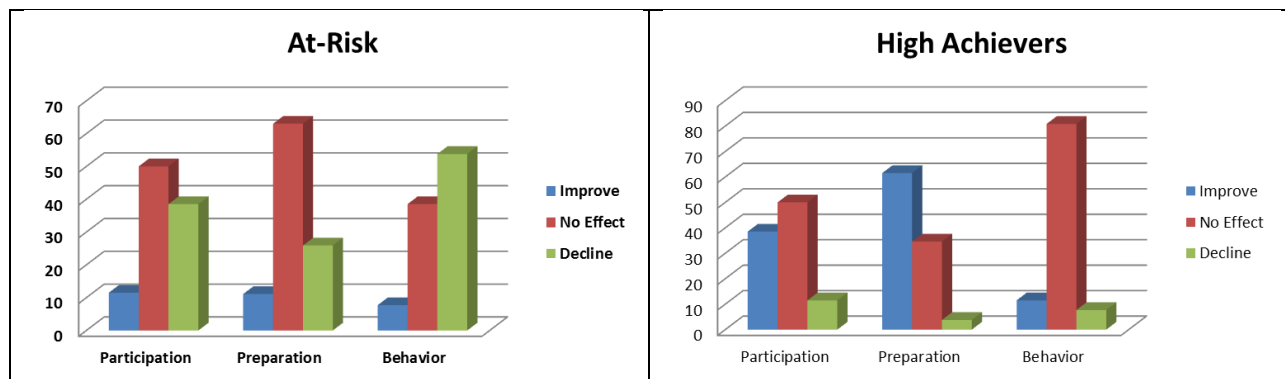
3.5 Facilitation of Communication Between Faculty and Students

The student computing program should facilitate communication between faculty and students. The faculty and students overwhelmingly agree that the laptop program helps to facilitate communication between the two groups. These results are illustrated in Figure 23.

Figure 23 – Perception of Effect of Laptop Program on Communication Between Faculty & Students
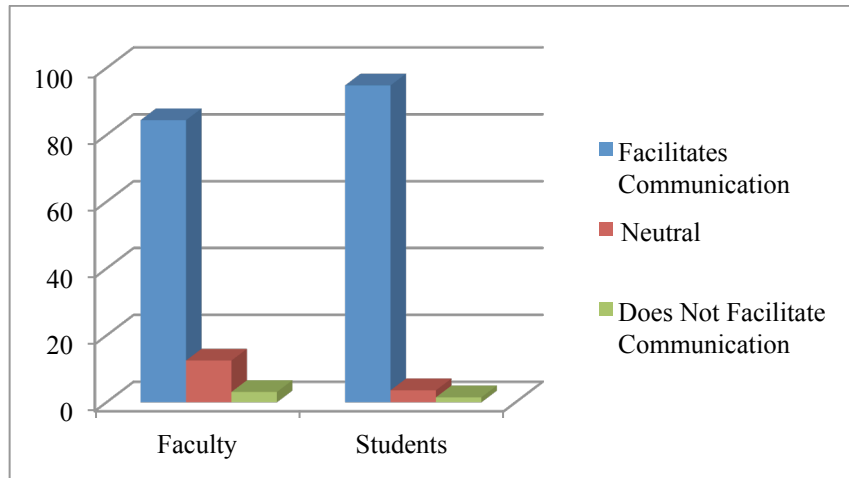
An additional indicator of the facilitation of communication between students and faculty is the frequency of use of technology-based communication tools. Students indicate that the faculty communicates with them regularly via Moodle, but does not make extensive use of other communication tools such as blogs, Ning, Skype, or Facebook. These results are shown in Figure 24.

Figure 24 – Students' Perceptions of Frequency of Use of Technology-Based Communication Tools
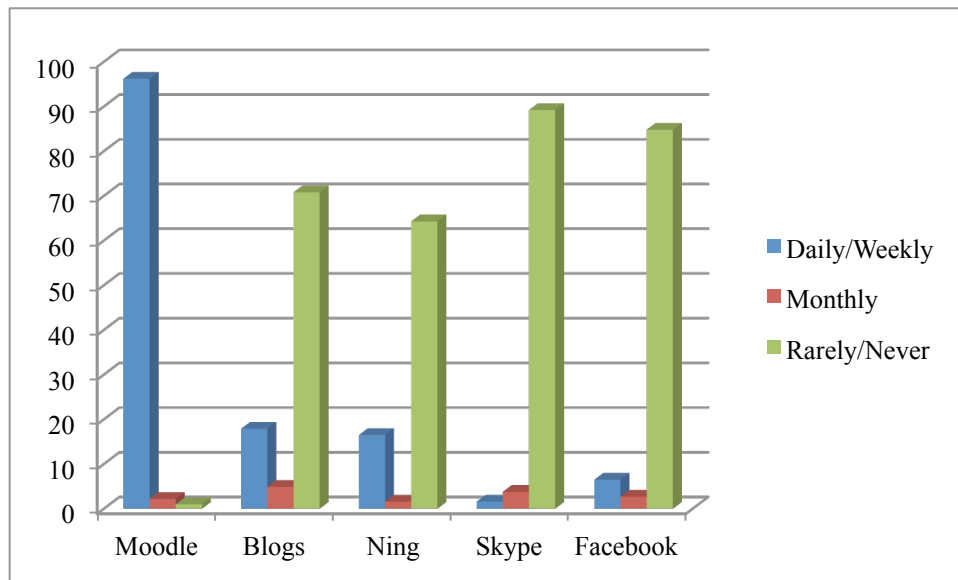
3.6 Facilitation of Marketing and Distinction From Other Schools in Area

The student computing program should help WXYZ to distinguish itself from other schools in the area and should facilitate marketing efforts. The two chosen indicators of success in this area are: the program as a factor in attracting new students, and the program in comparison to similar schools in the area.

3.6.1 Computing Program as Factor in Attracting New Students

The existence of a computing program has the potential to be a factor that attracts new students to the school. Current students and their parents disagreed about how big of a factor the existence of the program was in the decision-making process. More than half of current parents surveyed indicated that the existence of a computing program did not play any role in their decision to send their daughter to WXYZ. Alternatively, more than 75% of current students indicated that it played a role in their decision to attend. These results are illustrated in Figure 25.

Figure 25 – Existence of Computing Program as a Factor in Decision to Attend WXYZ
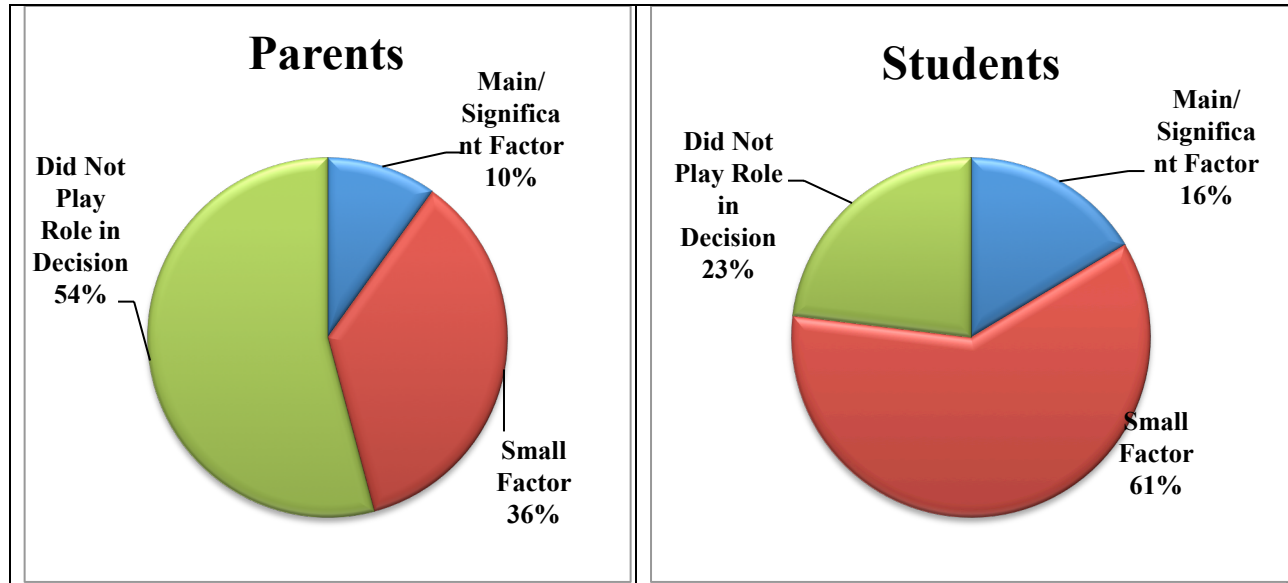
Parents of 8th grade girls can be divided into three groups: (1) those who will NOT send their daughter to WXYZ regardless of the technology program, (2) those who WILL send their daughter to WXYZ regardless of the technology program, and (3) those who will be swayed in their decision by the existence of an effective technology program. WXYZ will never attract the first group, and the results of the parent survey indicate that they have already captured the second group. Therefore, there appears to be an opportunity for WXYZ to attract members of the third group in the future.

3.6.2 Computing Program Compared to Other Schools

WXYZ was the first private school in the greater Detroit area to implement a one-to-one computing program in which the students were required to provide their own laptop. Sacred Heart Academy in Bloomfield Hills implemented a similar program several years ago, and has recently announced a decision to stay with a tablet computer rather than converting to an iPad program. University of Detroit Jesuit High School is rumored to be introducing a 1:1 iPad program during the 2013-2014 academic years. WXYZ's primary competitors, Marian and Ladywood, do not have strong technology programs, although Marian has mobile laptops on carts available for use.

### 3.7 Movement from Viability to Vitality

In 1998, the board of directors charged the administration with the goal of moving WXYZ from "viable to vital". The representative stakeholders who participated in this evaluation chose the following indicators as signs of vitality: enrollment trends, market share, faculty

attraction/retention, donation trends, relative tuition affordability, and changes in standardized test scores. See Section 3.1 for a discussion of test scores.

3.7.1 Enrollment Trends

All Catholic schools in the Archdiocese of Detroit have experienced challenges with enrollment during the past ten years. For the purpose of this report, WXYZ's enrollment was compared to three other all-girls Catholic schools in the area. These schools could be considered to be WXYZ's competitors for enrollment purposes. The schools include: Ladywood High School in Livonia, The Academy of the Sacred Heart in Bloomfield Hills, and Marian High School in Birmingham. Figure 26 shows the percentage change in enrollment for each school over the past ten years. If the line falls below the horizontal axis line it indicates that enrollment declined. The vertical line indicates the year in which the technology program was introduced. Sacred Heart, Ladywood, and Marian have experienced a drop in enrollment each year since the beginning of the recession in 2008. WXYZ is the only school to see an increase in enrollment during the past two years. There is no evidence, however, that the existence of the student computing program is responsible for this change.

Figure 26 - Percentage Change in Enrollment From Prior Year
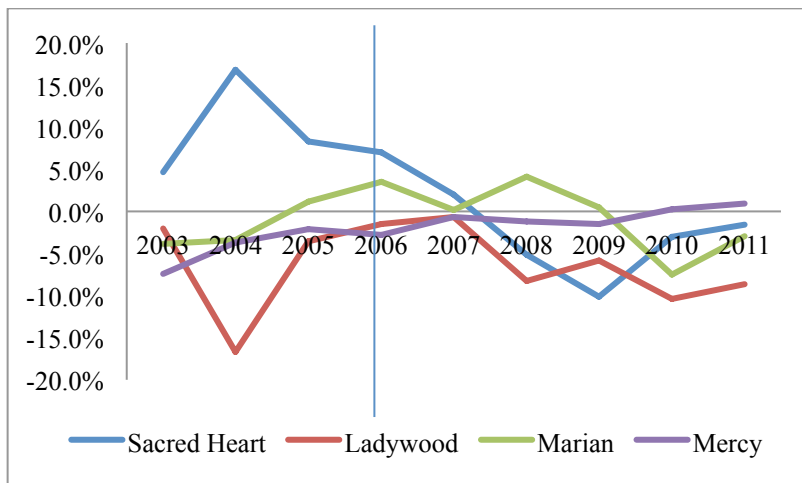
3.7.2 Market Share

Market share is the percentage of an "industry's" total size that is earned by a particular organization over a specified period of time. In this situation, it is calculated by taking WXYZ's enrollment and dividing it by the total enrollment for the four all-girl Catholic schools in the area (WXYZ, Marian, Sacred Heart, and Ladywood). WXYZ lost market share in the years prior to the introduction of the computing program, but has maintained or increased market share in the years since the program began. There is no evidence, however, that the existence of the student computing program is responsible for this change. Figure 27 illustrates this data.

Figure 27 – WXYZ's Percent of Market Share
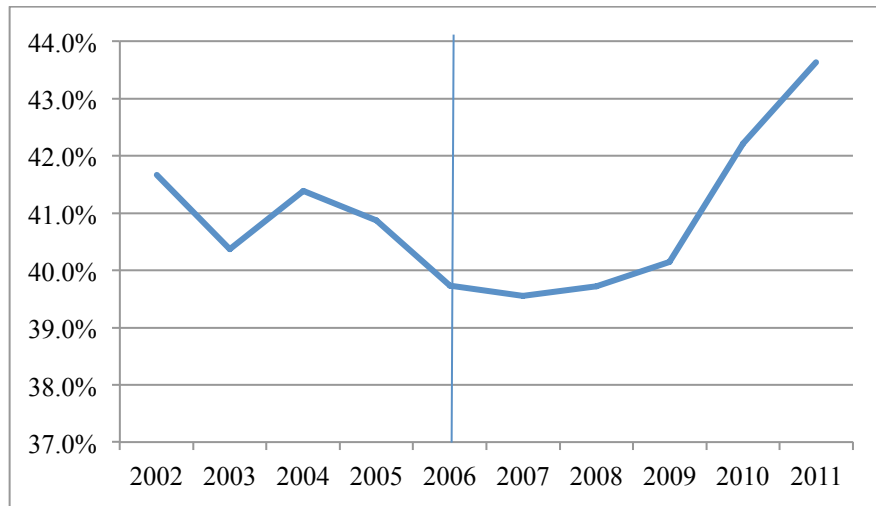
3.7.3 Faculty Attraction/Retention

Although the representative stakeholders who participated in this evaluation believed that data regarding the quality of faculty members and the school's ability to attract and retain outstanding faculty would be an indicator of the school moving from "viable to vital", the evaluator was unable to obtain reliable, measurable information about the faculty in prior years. Therefore, this indicator is not adequately addressed in this evaluation.

3.7.4 Donation/Endowment Trends

Donation trends are an indicator of a school's vitality. WXYZ's donation patterns do not lend themselves to easy interpretation, as seen in Figure 28. It is apparent however, that the technology program has not had a discernible impact. Further, parents who responded to the survey strongly indicated that the existence of an effective laptop/iPad program would not positively influence their donation decisions in the future, as seen in Figure 29. In fact, several parents indicated that they will donate *less* because of the expense associated with purchasing a laptop/iPad.

Figure 28 – Donation Trends for Past 10 Years
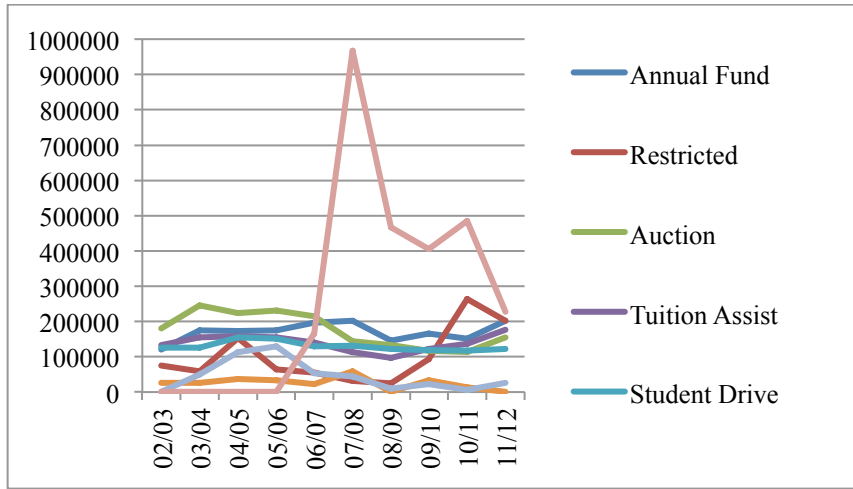
Figure 29 – Effect of Technology Program on Parents' Donation Decisions
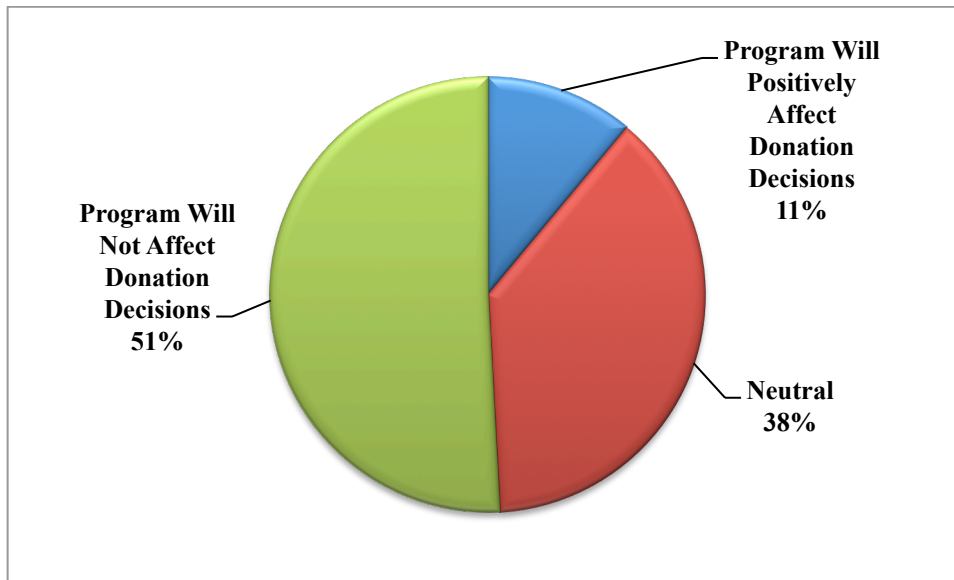
3.7.5 Tuition Affordability

Maintaining an affordable tuition rate in the face of economic challenges is an additional indicator of a vital school. Figure 30 illustrates the relative tuition rates for WXYZ and its competitor all-girls schools as well as the tuition rates of two nearby all-boys Catholic High Schools for the 2012-2013 school year. WXYZ's tuition rate is in line with other schools in the area. WXYZ parents' responses were evenly split when they were asked whether they thought the total cost of sending their daughter to WXYZ, including tuition and the cost of the laptop/iPad, is reasonable, as seen in Figure 31. However, it is clear that parents did not think the cost of the HP laptop was reasonable. Their responses included many emotionally charged words, such as "nightmare" and "outrageous". The change to iPads was welcomed by many parents as a more reasonable alternative in terms of cost.

Figure 30 – Relative Tuition Rates for Catholic School in the Area
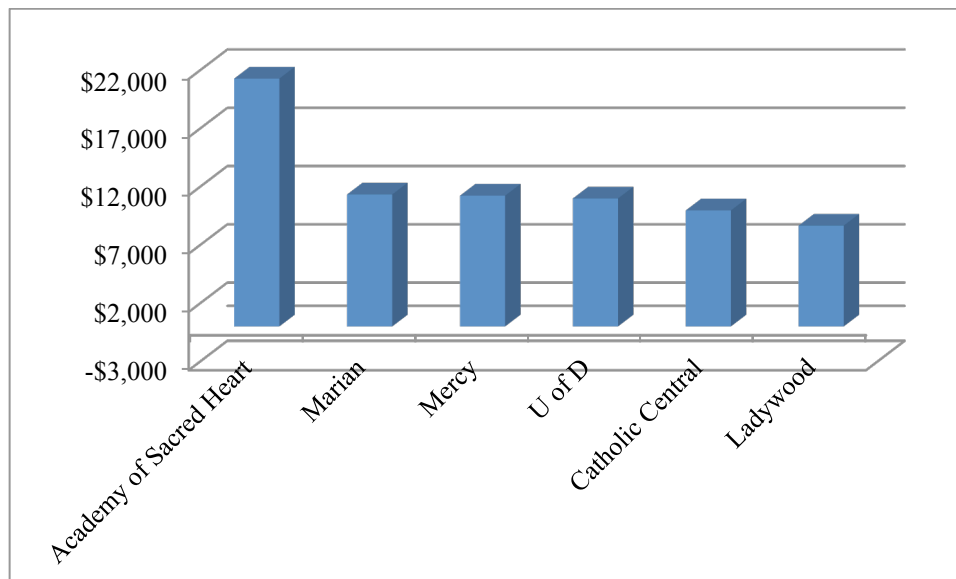
Figure 31 – Parents' Perception of Reasonableness of Tuition
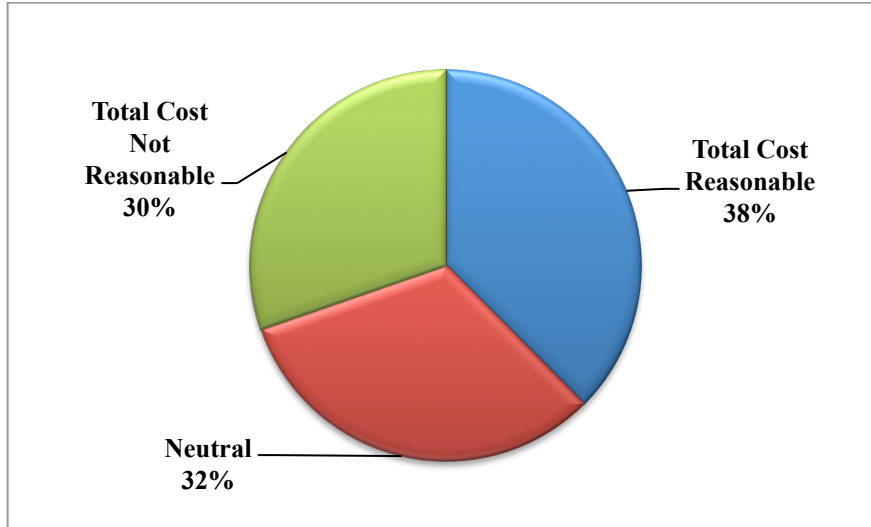
3.8 Incorporation of Global Shift in Learning

In recent years there has been a shift among educators from a mindset which describes learning as an assimilation of facts towards a description of learning as "knowledge creation and use by the individual". This shift re-casts the student from the role of passive recipient of information to a more active role with personal responsibility. Encouraging students to embrace this active role is a good way to prepare them for lifelong learning in the 21st century. The student computing program should facilitate the faculty's efforts in this regard.

3.8.1 Implementation of Challenge Based Learning

According to the Apple Corporation, challenge based learning (CBL) is a multidisciplinary approach to teaching and learning that requires students to use the technology that is part of their everyday lives to solve real-world problems. Research in the field of education suggests that asking students to try to solve ill-structured, real world problems is an effective way to help them to develop critical thinking skills (See section 3.1.6). In a challenge based learning project, students are presented with a big-picture problem and then asked to draw on prior learning, acquire new knowledge, and develop a creative solution. The projects are collaborative, hands on, and require students to work together, with their teachers, and with experts. WXYZ's administration strongly encouraged faculty to include a CBL project in their curriculum during the 2011-2012 school year. More than 80% of the faculty indicated that they participated in a CBL project, and 93% of the projects included a technology-based component. The Math faculty indicated the lowest participation rate in CBLs. The English faculty indicated the highest participation rate in one or more CBLs. 80% of students also indicated that they participated in a CBL project. More than 80% of students indicated that there was a technology component in their CBL projects (17% were unsure, 1.7% did not include a technology component).

3.8.2 Faculty Professional Development

Professional development contributes to the faculty's ability to incorporate the global shift in learning into their curriculum. See Section 3.4.1 and Figure 21 for a description of faculty professional development.

3.8.3 Faculty Use of Technology-Based Teaching Methods

In the 2011 internal faculty survey, faculty members expressed enthusiasm for a wide variety of technology-based teaching tools. However, in the more recent survey faculty members indicated that they have not incorporated a wide variety of tools into their teaching, as seen in Figure 32. This is a major finding in this evaluation and an area of concern. This should be an area of focus in the future.

Figure 32 – Frequency of Use of Technology-Based Teaching Tools
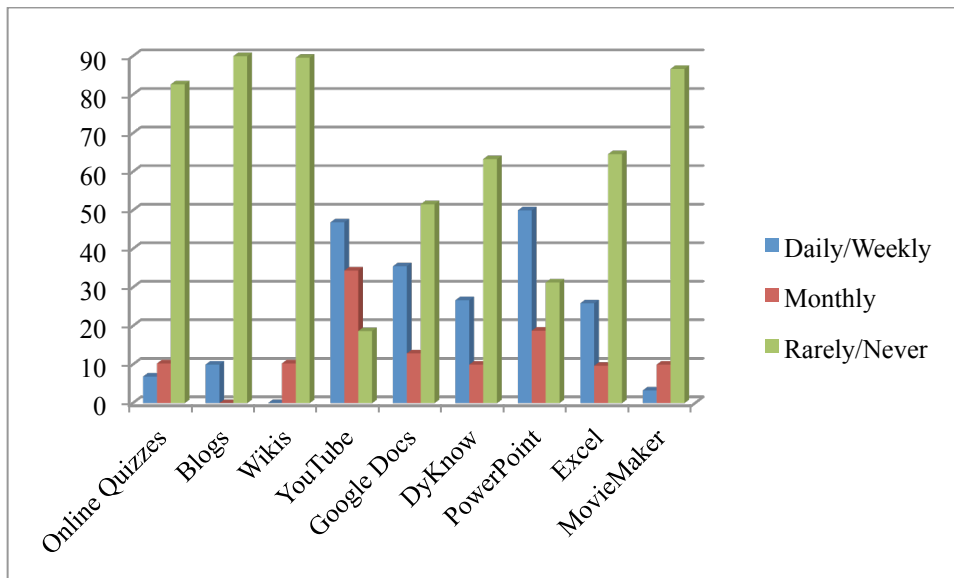
3.8.4 Technology Integration in All Areas of Curriculum

In the recent faculty survey it was not surprising to find that integration of technology varied by department. For example, the English and Social Studies departments were the only ones to use blogs and Ning. The Science department uses Google Docs more frequently than other departments. Interestingly, the Language department reported the use of more *types* of technology based tools than any other department. However, as mentioned above, the use of most of these technology-based teaching tools is not widespread, and leads to the conclusion that technology is not widely integrated into all areas of the curriculum. The administration encourages but does not require that faculty members integrate technology into their classes.

3.9 Extension of Learning Environment Boundaries

The overwhelming perception on the part of faculty and students is that the computing program has facilitated communication between the two groups (see Section 3.5). This suggests that the learning environment has been extended beyond 8 a.m. – 3 p.m. In addition, the widespread use of Moodle (more than 80% of faculty uses it daily or weekly) as a learning management system is evidence that the computing program is facilitating the extension of classroom boundaries. However, the relatively limited use of technology-based teaching methods and technology integration into all areas of the curriculum is a limiting factor for this desired result.

3.10 Integration of Values Into Teaching and Learning

One of the goals identified in the 1998 Technology Vision Statement is that technology should promote the integration of values into all aspects of the teaching/learning process. One way for this to occur is to ensure that technology is incorporated into classes that specifically address the development of values. The faculty members from the Religious Studies department who took the recent faculty survey identified that they use all of the following technology based teaching tools at least monthly: Moodle, online quizzes, wikis, Audacity, YouTube, Google Docs, Skype, DyKnow, PowerPoint, Excel, and MovieMaker.

3.10.1 Movement Towards Green Technologies

The representative stakeholders who participated in the planning of this evaluation identified the movement towards green technologies as an additional indicator of the promotion of values into teaching and learning. As part of the current transition to an iPad program, the administration has taken several steps move towards a paperless school. First, they have negotiated terms with the online bookstore in order to put ebooks on a level playing field with their paper counterparts. In addition, they are encouraging the academic departments within the school to adopt textbooks with digital versions. Finally, the administration is requiring teachers to allow and accommodate students who use digital texts.

3.10.2 Technology Acceptable Use and Cyberbullying Policies

The WXYZ Agenda Book that is distributed to each student contains an extensive technology-related Student Acceptable Use Policy that clearly identifies appropriate boundaries for access to and usage of all computing systems related to the school. In addition, the Agenda Book has a thorough section devoted to online safety and the prevention and handling of cyberbullying. These policies are reviewed with students at the beginning of each year.

3.11 Educating Women Who Make a Difference

At the strategic level, W seeks to contribute to society by adhering to their motto "Educating Women Who Make a Difference". While there are countless ways to make a difference, for the purposes of this evaluation, the representative stakeholders identified three measurable indicators of growth in this area: the number of scholarships awarded to graduates, the quality of colleges

that students attend, and the number of students who decide to pursue majors/careers in Science, Technology, Engineering, or Math (STEM).

3.11.1 Number of Scholarships Awarded to Graduates

An effective student computing program should lead to a stronger curriculum, which should, in turn, lead to an increase in the number of scholarships awarded to graduates. Students who receive scholarships have the potential to eventually "make a difference" in the world. WXYZ has experienced a trend towards a higher number of scholarships per student in the past 9 years, as seen in Figure 33. However, there is no proof that the student computing program is responsible for this trend.

Figure 33 – Average # of College Scholarships Received per Student
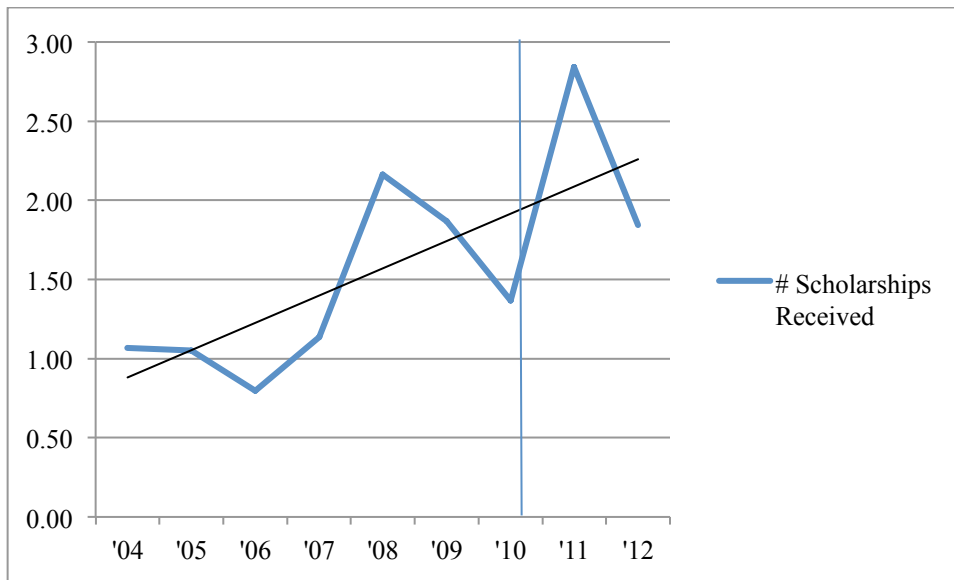
3.11.2 Quality of Colleges Attended

By pursuing higher education at outstanding colleges and universities, WXYZ students enhance their potential to become "Women Who Make a Difference". By examining the quality of the universities that WXYZ students attend, it is possible to evaluate this indicator. There are many college ranking systems available to use for this study. The evaluator chose the U.S. News and World Report College Ranking Report for 2012 because it is widely used and respected. US News and World Report ranks colleges in several categories, including: national schools, national liberal arts colleges, and regional schools. The top 200 schools are listed in each category. Using this ranking system, the evaluator assigned a number to every graduate for the past seven years and calculated an average quality number for each class. For example, if a student attended Harvard, she received a rank of "1" because US News and World Report ranks Harvard as the best school in the country. The University of Michigan is ranked at #28 while

Michigan State is ranked at #71. Regional schools, such as Albion and Adrian were ranked from 200-400. Schools that were not ranked in the top 200 national schools or the top 200 regional schools, including Central Michigan University and Wayne State University, received a ranking of 400. For a complete description of the method used to calculate the scores, please contact the evaluator. The results of the ranking appear in Figure 33 and show the average college ranking for the past 7 years. The vertical line indicates the graduation year of the first students who participated in the student computing program. For the purposes of this indicator, a low number is desirable (for example, if every girl in the class attended Harvard, the class would have an average score of 1. If every girl attended Wayne State, the class would have an average score of 400). The horizontal axis is reversed in Figure 34 to show a positive trend. WXYZ has seen an improvement in the quality of colleges chosen by graduates in the past 7 years. However, there is no proof that the student computing program is responsible for this trend.

Figure 34 – Average Rank of Colleges Attended by WXYZ Students
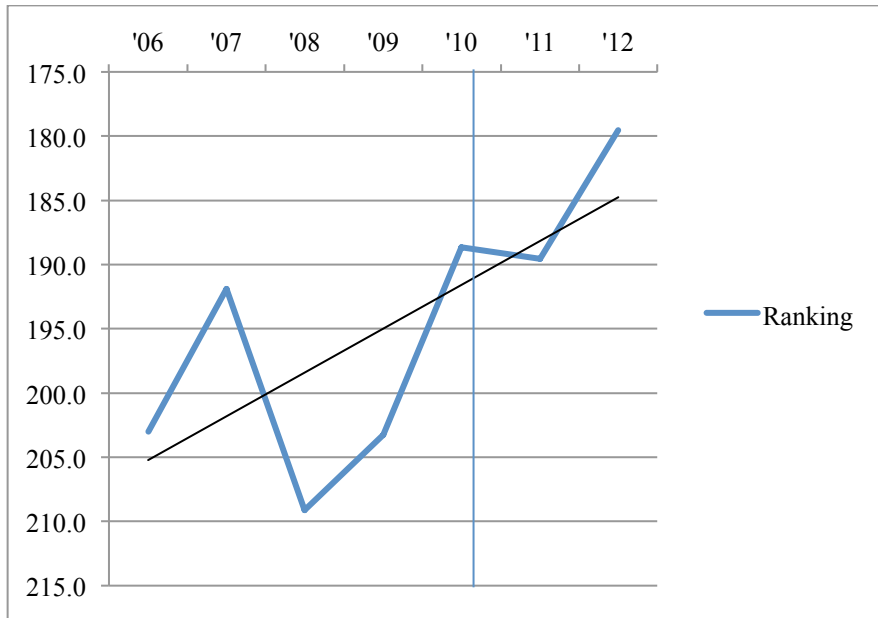
3.11.3 Number of Students Who Pursue STEM Majors/Careers

The state of Michigan has experienced a severe economic downturn during the past few years. As a result, there has been a net reduction in the state's population as workers leave in search of employment elsewhere. One of the stated goals of organizations such as the Michigan Council of Women in Technology is to reverse this trend by offering educational opportunities to young people that will strengthen their skills in Science, Technology, Engineering, and Math, which are known as the STEM fields. These fields are at the heart of Michigan's industrial and manufacturing base. An effective student computing program should provide students with a

solid foundation in these areas, and potentially encourage students to pursue higher education in these fields. Figure 35 illustrates the percentage of WXYZ students who reported on the PSAT that they planned to pursue STEM related education. The positive trend shown in Figure 34 is encouraging. However, there is no corresponding trend towards STEM-related scholarships. In fact, STEM scholarships have declined overall during the same time period, as seen in Figure 36. Although the existence of an effective student computing program seems likely to encourage a positive attitude towards technology-based fields, there is no proof that the program is responsible for the increase in STEM majors.

Figure 35 - % Students Pursuing STEM Major
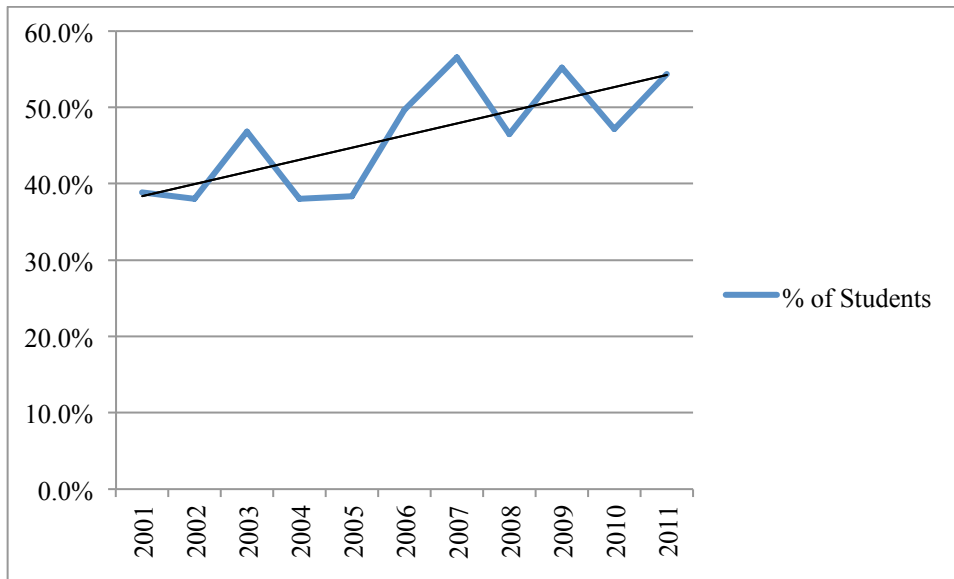
Figure 36 – STEM-Related Scholarships
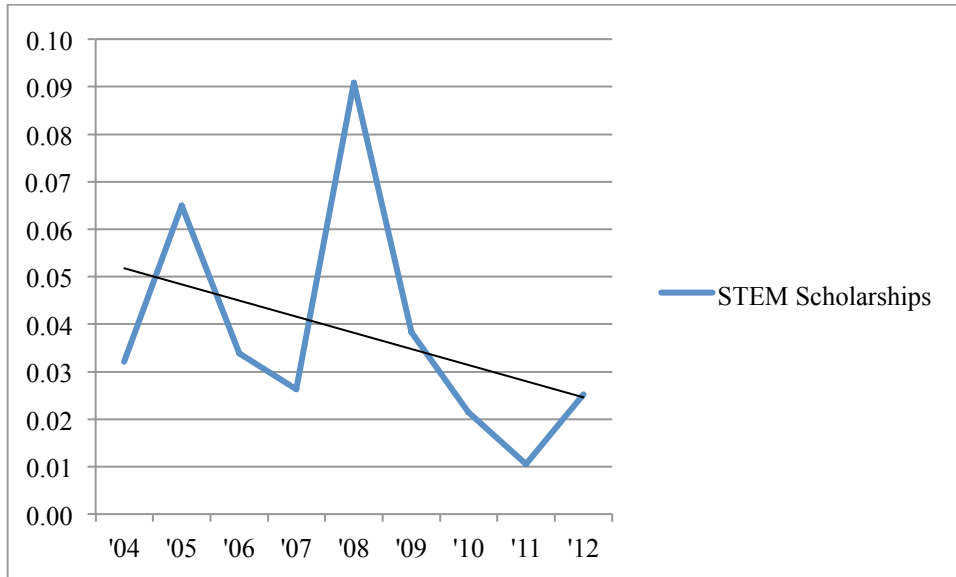
4. Conclusions and Recommendations

4.1 Conclusions

This evaluation tried to answer eleven questions. Here are the conclusions that have been reached for each question as well as the corresponding recommendations for action.

1. Has the program helped students to develop college and 21st Century skills?

WXYZ students have seen an increase in composite ACT scores as well as individual scores in Math, English, and Science. Recent graduates believe that the student computing program improved their written communication skills, as well as their ability to access and analyze information. They also believe that the program helped them learn how to organize information and helped them develop critical thinking skills.

Unfortunately, WXYZ has seen a decrease in PSAT Writing Skills scores during the past ten years. In addition, ACT Reading scores have been flat or declined slightly. While the technology program cannot be directly credited with the increase in composite scores, or directly blamed for the decrease in Writing Skills scores, WXYZ should focus technology integration efforts to support these weak areas. There are a wide variety of technology-based tools that are designed to improve writing skills, and this is an opportunity for WXYZ to use the technology program to target a specific goal.

2. Has the program increased students' competence and confidence in the creation and use of information?

Current students and recent graduates believe that the student computing program has improved their competence and confidence in the use of technology overall.

The only areas that need improvement are the creation of spreadsheets and video/digital editing. Students reported that they rarely have the opportunity to use these tools. WXYZ should focus technology integration efforts to help faculty find ways to work opportunities to use these resources into the curriculum.

3. Has the program enriched the classroom experience and increased motivation by making learning more enjoyable and engaging students in their own knowledge creation and use?

Current students overwhelmingly believe that laptops make school work easier to do and more interesting. They also believe that the laptop program has improved the quality of their work. However, the faculty is not as enthusiastic. While a high percentage of faculty members believe that the program has improved students' level of preparedness, engagement, and quality of work, they also believe that the program has caused a decline in participation and behavior. This is a key area of concern. The administration and faculty should make it a priority to address and reverse this finding.

4. Has the program helped the faculty to implement the curriculum more effectively and facilitated the individualization of instruction?

The faculty reported that they are satisfied, overall, with the level of professional development that they have received as the school makes the transition to iPads. However, the English, Language, Performing Arts, and Math faculty expressed concern that they have not received enough professional development to effectively integrate technology into their curriculum.

Faculty reports good results for high achieving students as a result of the program. However, the faculty believes low achieving/at-risk students have experienced a decline in participation and behavior. This is also a key area of concern and should be a priority as the school moves forward.

5. Has the program facilitated communication between students and faculty?

Students and faculty overwhelming reported that the program has facilitated communication between the two groups. This communication appears to occur primarily through Moodle. While this is a very strong positive result of the program, future technology integration initiatives should include the exploration of addition technology-based communication tools.

6. Has the program helped WXYZ distinguish itself from other schools in the area and has it facilitated marketing?

Parents of current students overwhelmingly responded that the technology program was not a factor in their decision to send their daughter to WXYZ. However, it is likely that there are parents out there who would be swayed in their choice of high school by the existence of an effective technology program. This evaluation uncovered a lot of good news about WXYZ, including:

- The school has seen almost a full point increase in its Composite ACT scores during the past ten years.
- The school has seen increases in several ACT subject scores, including Math, English, and Science
- The school has seen an increase in the PSAT critical reading score
- WXYZ is the only school among its competitors to see an increase in enrollment in the past several years
- WXYZ's percentage decrease in enrollment during the recent recession was less than its competitors
- The school has seen an increase in market share in the past 5 years
- The school has seen an increase in the number of college scholarships received per student during the past nine years
- The school's overall donation levels remained strong throughout the recent recession
- Students have attended increasingly competitive colleges over the past ten years
- The school has seen an increase in the number of students pursuing STEM-related fields of study
- Recent graduates who participated in the student computing program and then went on to college believe that the program helped them to improve their writing skills
- 95% of recent graduates believe that the program improved their ability to access information
- Both recent graduates and current students believe that the program has increased their ability to analyze information
- Recent graduates believe that the program improved their ability to organize information
- Recent graduates feel like they have the same or better critical thinking skills as their college peers
- 96% of recent graduates feel that they have the same or better ability to use technology as their peers in college
- The majority of current students believe that the program has improved the quality of their work
- Recent graduates feel that the program improved their competence in the creation of presentations and spreadsheets, and in the use of databases, social media and video/digital editing

- Current students feel that the program has made their schoolwork more interesting and easier
- Faculty and current students overwhelmingly believe that the program has improved the communication between the two groups.

Some of this good news is attributable to the technology program and some is not. However, it is all available for use in the marketing of the school to potential families. WXYZ's technology program remains at the forefront of programs in the area. The school has an excellent opportunity to spread the good news.

7. Has the program helped WXYZ move from "viable to vital"?

WXYZ shows many signs of having answered the board of directors' call to move the school from "viable to vital". Enrollment trends, market share, donation trends, and relative tuition affordability are all positive, even in the face of the recent recession. It is unclear whether the technology program is responsible for this movement, but the fact remains that the school shows signs of vitality.

8. Has the program incorporated the global shift in learning towards the concept of "knowledge creation and use by the individual"?

More than 80% of faculty and students indicated that they participated in a challenge-based learning project during the 2011-2012 school year. In a challenge-based project, students are presented with a problem and asked to develop a creative solution using prior learning, teamwork, and new information. Almost all of the projects included a technology component. Recent educational research indicates that these types of projects help students to reach a deeper understanding of a subject by allowing them to create their own knowledge during the process. Twenty percent of teachers did not participate in a challenge-based project even though the school administration indicated to them that it was required. Individual teachers stated that there were no repercussions for failing to participate in a project. WXYZ should have mandatory technology integration requirements for all faculty members, and failure to incorporate technology into the curriculum should be reflected in the annual performance review process.

Widespread incorporation of technology based teaching methods is another indicator of the incorporation of a shift in learning towards individual knowledge creation and use. While the faculty expressed enthusiasm for a wide variety of technology-based teaching tools, they indicated that they actually use relatively few of them in the classroom (see Figure 32). Future professional development and technology integration efforts should assist the faculty with this issue.

9. Has the program extended the boundaries of the learning environment beyond the school building and 8 a.m. – 3 p.m.?

The overwhelming perception on the part of faculty and students is that the computing program has facilitated communication between the two groups. The primary means of communication is the learning management system called Moodle, which is available to students and faculty members 24 hours a day. This suggests that the learning environment has been extended beyond 8 a.m. – 3 p.m.

10. Has the program promoted the integration of values into teaching and learning?

It is difficult to measure the effect of the technology program on the integration of values. However, there are several signs that WXYZ is factoring values into the use of technology within the school. Faculty members in the Religious Studies department indicated that they use a wide variety of technology based teaching tools in their classes. The school continues to pursue green technologies, such as eTextbooks. The school has a comprehensive policy about the acceptable use of technology and about cyberbullying. Students receive training about the safe use of technology at the beginning of each school year.

11. Has the program helped WXYZ to educate women who make a difference and foster the spiritual, intellectual, moral, physical, and cultural development of its students?

While there are countless ways to make a difference in the world, the stakeholders in this evaluation chose three indicators of this goal: the number of scholarships awarded to graduates, the quality of colleges attended, and the number of students who pursue STEM-related studies/careers. WXYZ has seen positive results in all three of these areas during the past several years. The number of scholarships received by graduates each year has almost doubled. Students are attending increasingly challenging colleges, and more than half of students currently indicate that they plan to pursue studies in Science, Technology, Engineering or Math. While the technology program cannot be directly credited with these positive results, it appears to be part of an increasing culture of achievement within the school.

4.2 Recommendations

Based on the results of this evaluation, the evaluator makes the following four recommendations:

1. Market the School's Successes.

This evaluation has uncovered many positive trends at WXYZ. The school should use this information to market itself to potential families. In particular, the increase in ACT scores, enrollment, market share, the number of college scholarships per student, and the number of students pursuing STEM-related careers/majors are significant signs of vitality and should be a focus of marketing efforts.

2. Address Faculty Concerns About Decline in Participation and Behavior

The faculty's perception that the technology program has caused a decline in class participation and behavior is a major cause of concern. The faculty and administration should address this concern directly. Together, the groups should identify what works and what doesn't work with students, and focus on improving these results for the benefit of faculty and students.

3. Support At-Risk Students

The technology program should provide the faculty with an opportunity to help at-risk students reach new levels of achievement. However, the results of the evaluation indicate that at-risk students are experiencing more harmful effects from the program than their high achieving classmates. The faculty and administration should clearly articulate the support of at-risk students as a goal for this academic year. The two groups should work collaboratively to identify specific steps that can be taken to address this concern and reverse this finding.

4. Technology Integration – Require It, Support It

With the transition to iPads, the administration has done an admirable job of providing professional development opportunities for the faculty. In general, the faculty is enthusiastic about the technology program and the possibilities that exist for them to use technology to facilitate learning among their students. The reality, however, is that their actual incorporation of technology is limited (see Figure 32). After seven years of this program, faculty members have had time to acclimate themselves to the use of technology and should be required to incorporate it into their curriculum. Technology integration efforts should be considered during the annual performance review process.

If technology integration is going to be required and measured, the school administration needs to continue to support the faculty in their technology integration efforts. Recent professional development has been a positive step, but it would be effective at this point to provide faculty members with individual attention. Since Mr. Baker and the technology support staff are already fully occupied, the school should consider hiring a technology integration specialist on a part-time or hourly basis to work with individual faculty members to find ways to effectively incorporate technology-based teaching methods into the curriculum. This person could recommend proven technology-based tools to individual teachers and support their implementation. The first priority for a technology integration specialist should be to find tools to improve writing skills, in an effort to reverse the disappointing trend in PSAT writing skills scores.

The technology program has been a success at WXYZ. The culture of the school is increasingly one of achievement and openness to new ideas. The school shows many signs of moving towards vitality. With additional attention to these areas of concern, the school has an opportunity to achieve great things.

APPENDIX K – METAEVALUATION RESULTS BY PARTICIPATING GROUP

Standard		Stake Holders	Meta Evaluator	Evaluator
Utility				
U1. Evaluator Credibility	U1.1 Engage evaluators who possess the needed knowledge, skills, experience, and professional credentials	X	X	X
	U1.2 Engage evaluators whose evaluation qualifications, communication skills, and methodological approach are a good fit to the stakeholders' situation and needs	X	X	X
	U1.3 Engage evaluators who are appropriately sensitive and responsive to issues of gender, socioeconomic status, race, language, and culture	X	X	X
	U1.4 Engage evaluators who build good working relationships, and listen, observe, clarify, and attend appropriately to stakeholders' criticisms and suggestions	X	X	X
	U1.5 Engage evaluators who have a record of keeping evaluations moving forward while effectively addressing evaluation user's information needs	X	X	
	U1.6 Give stakeholders information on the evaluation plan's technical quality and practicality, e.g., as assessed by an independent evaluation expert	X	X	X
U1 – Evaluator Credibility - Total		**6**	**6**	**5**
U2. Attention to Stake-holders	U2.1 Clearly identify and arrange for ongoing interaction with the evaluation client	X	X	X
	U2.2 Identify and arrange for appropriate exchange with the other right-to-know audiences, including, among others, the program's authority figures, implementers, beneficiaries, and funders	X	X	X
	U2.3 Search out and invite input from groups or communities whose perspectives are typically excluded, especially stakeholders who might be hindered by the evaluation	X	X	X
	U2.4 Help stakeholders understand the evaluation's boundaries and purposes and engage them to uncover assumptions, interests, values, behaviors, and concerns regarding the program	X	X	X
	U2.5 Determine how stakeholders intend to use the evaluation's findings	X	X	X
	U2.6 Involve and inform stakeholders about the evaluation's progress and findings throughout the process, as appropriate	X	X	X
U2 – Attention to Stakeholders - Total		**6**	**6**	**6**

Standard		Stake Holders	Meta Evaluator	Evaluator
Negotiated Purpose	U3.2 Engage the client and stakeholders to weigh stated evaluation purposes – e.g. against their perceptions of dilemmas, quandaries, and desired evaluation outcomes – and to embrace evaluation's bottom line goal of assessing value, e.g., a program's worth, merit, or significance	X	X	X
	U3.3 Help the client group consider possible alternative evaluation purposes, e.g., program planning, development, management, and improvement: program documentation and accountability; and judging the program's quality, impacts, and worth		X	X
	U3.4 Engage the client to clarify and prioritize the evaluation's purposes using appropriate tools such as needs assessments and logic models	X	X	X
	U3.5 Provide for engaging the client group periodically to revisit and, as appropriate, update the evaluation's purpose	X	X	X
	U3.6 Assure that initial and updated evaluation purposes are communicated to the full range of stakeholders	X	X	X
U3 – Negotiated Purpose - Total		**5**	**6**	**6**
U4. Explicit Values	U4.1 Make clear the evaluator's commitment to certain, relevant values, e.g., an evaluation's utility, feasibility, propriety, accuracy, and accountability and a program's equity, fairness, excellence, effectiveness, safety, efficiency, fiscal accountability, legality, and freedom from fraud, waste, and abuse	X	X	X
	U4.2 Engage the client and program stakeholders in an effective process of values clarification, which may include examining the needs of targeted program beneficiaries, the basis for program goals, and the rationale for defined evaluation purposes	X	X	X
	U4.3 Assist the client group to air and discuss their common and discrepant vies of what values and purposes should guide the program evaluation	X	X	X
	U4.4 Acknowledge and show respect for stakeholders' possibly diverse perspectives on value matters, e.g., by assisting them to seek consensus or at least reach an accommodation regarding possible alternative interpretations of findings against different values	X	X	X
	U4.5 Clarify the values that will undergird the evaluation, taking account of client, stakeholder, and evaluator positions on this matter	X	X	X
	U4.6 Act to ensure that the client and full range of stakeholders understand and respect the	X	X	X

Standard		Stake Hold ers	Meta Eval uator	Eval uator
	values that will guide the collection, analysis, and interpretation of the evaluation's information			
U4 – Explicit Values - Total		**6**	**6**	**6**
U5. Relevant Information	U5.1 Interview stakeholders to determine their different perspectives, information needs, and views of what constitutes credible, acceptable information	X	X	X
	U5.2 Plan to obtain sufficient information to address the client group's most important information needs	X	X	X
	U5.3Assess and adapt the information collection plan to assure adequate scope for assessing the program's value, e.g., its worth, merit, or significance	X	X	X
	U5.4 Assure that the obtained information will address and keep within the boundaries of the evaluation's stated purposes and key questions	X	X	X
	U5.5 Allocate time and resources to collecting different parts of the needed information in consideration of their differential importance	X	X	X
	U5.6 Allow flexibility during the evaluation process for revising the information collection plan pursuant to emergence of new, legitimate information needs			X
U5 – Relevant Information - Total		**5**	**5**	**6**
U6. Meaningful Processes and Products	U6.1 Budget evaluation time and resources to allow for meaningful exchange with stakeholders throughout the evaluation process	X	X	X
	U6.2 Engage the full ranges of stakeholders to assess the original evaluation plan's meaningfulness for their intended uses	X	X	X
	U6.3 During the evaluation process, regularly visit with stakeholders to assess their evaluation needs and expectations, also, as appropriate, to obtain their assistance in executing the evaluation plan	X	X	X
	U6.4 Regularly obtain stakeholders' reactions to the meaningfulness of evaluation procedures and processes	X	X	X
	U6.5 Invite stakeholders to react to and discuss the accuracy, clarity, and meaningfulness of evaluation reports	X	X	X
	U6.6 As appropriate, adapt evaluation procedures, processes, and reports to assure that they meaningfully address stakeholder needs	X	X	X

202

Standard		Stake Holders	Meta Evaluator	Evaluator
U6 - Meaningful Processes and Products - Total		**6**	**6**	**6**
U7. Timeliness and Appropriate Communication and Reporting	U7.1 Plan to deliver evaluation feedback pursuant to the client group's projection of when they will need reports, but allow flexibility for responding to changes in the program's timeline and needs	X	X	X
	U7.2 Plan, as appropriate, to give stakeholders access to important information as it emerges	X		
	U7.3 Employ reporting formats and media that accommodate the characteristics and serve the needs of the different audiences	X	X	X
	U7.4 Determine how much technical detail to report by identifying and taking account of the audience's technical background and expectations	X	X	X
	U7.5 Plan and budget evaluation follow-up activities so that the evaluator can assist the client group to interpret and make effective use of the final evaluation report	X		
	U7.6 Pursuant to the above checkpoints, formalize expectations for communicating and reporting to the sponsor and stakeholders in the evaluation contract	X		X
U7 - Timeliness and Appropriate Communication and Reporting - Total		**6**	**3**	**4**
U8. Concern for Consequences and Influence	U8.1 Identify the stakeholders' formal and informal communication mechanisms that connect stakeholders and, as appropriate, channel evaluation findings through these mechanisms	X	X	X
	U8.2 Be vigilant and proactive in identifying and appropriately communicating with stakeholders who appear to be sabotaging the evaluation and, as necessary, counteract the sabotage			
	U8.3 Plan to meet, as appropriate, with stakeholders to help them apply findings in ways that are logical, meaningful, ethical, effective, and transparent	X	X	X
	U8.4 In discussing evaluation findings with the client group stress the importance of applying the findings in accordance with the evaluation's negotiated purposes	X		X
	U8.5 Be vigilant to identify, prevent, or appropriately address any misuses of evaluation findings	X		
	U8.6 Follow up evaluation reports to determine if and how stakeholders applied the findings	NA		X

Standard		Stake Holders	Meta Evaluator	Evaluator
U8 - Concern for Consequences - Total		**4**	**2**	**4**
Feasibility				
F1. Project Management	F1.1 Ground management of the evaluation in knowledge of the stakeholders' environment and needs and the evaluation's purpose	X	X	X
	F1.2 Prepare a formal management plan including, e.g., the evaluation's goals, procedures, assignments, communication, reporting, schedule, budget, monitoring arrangements, risk management arrangements, and accounting procedures	X		X
	F1.3 Recruit evaluation staff members who collectively have knowledge, skills, and experience required to execute, explain, monitor, and maintain rigor, viability, and credibility in the evaluation process	X		X
	F1.4 Involve and regularly inform an appropriate range of stakeholders	X	X	X
	F1.5 Systematically oversee and document the evaluation's activities and expenditures	X	X	X
	F1.6 Periodically review the evaluation's progress and, as appropriate, update the evaluation plan and procedures	X	X	X
F1 – Project Management - Total		**6**	**4**	**6**
F2. Practical Procedures	F2.1 Assess and confirm the program's evaluability before deciding to proceed with the evaluation	X		X
	F2.2 Employ procedures that fit well within the program and its environment	X	X	X
	F2.3 Assure that the selected procedures take account of and equitably accommodate the characteristics and needs of diverse stakeholders	X	X	X
	F2.4 Obtain relevant insider knowledge and incorporate it into the data collection process		X	X
	F2.5 Make efficient use of existing information and avoid needless duplication in collecting data	X	X	X
	F2.6 Conduct the evaluation so as to minimize disruption to the program	X	X	X
F2 – Practical Procedures - Total		**5**	**5**	**6**
F3. Contextual Viability	F3.1 Investigate the program's cultural, political, and economic contexts by reviewing such items as the program's funding proposal, budget documents, organizational charts, reports, and news media accounts and by interviewing such stakeholders as the program's funder, policy board members, director, staff, recipients, and area residents	X	X	X

Standard		Stake Holders	Meta Evaluator	Evaluator
	F3.2 Take into account the interests and needs of stakeholders in the process of designing, contracting for, and staffing the evaluation	X	X	X
	F3.3 Enlist stakeholder and interest group support through such means as regular exchange with a review panel composed of a representative group of stakeholders	X	X	X
	F3.4 Practice even-handedness and responsiveness in relating to all stakeholders, e.g., in the composition of focus groups	X	X	X
	F3.5 Avert or identify and counteract attempts to bias or misapply the findings			
	F3.6 Provide appropriate mechanisms for stakeholders to remain informed about the evaluation's progress and findings, such as an evaluation project website, an evaluation newsletter, targeted reports, and a telephone response line	X	X	
F3 – Contextual Viability - Total		**5**	**5**	**4**
F4. Resource Use	F4.1 Negotiate a budget--ensuring that the contracted evaluation work can be completed efficiently and effectively—to include the needed funds and the necessary in-kind support and cooperation of program personnel			X
	F4.2 Balance effectiveness and efficiency in resource use to help ensure that the evaluation will be worth its costs and that sponsors will get their money's worth		X	
	F4.3 Use resources carefully with as little waste as possible	X	X	X
	F4.4 Utilize existing data, systems, and services when they are well aligned with the evaluation's purposes	X	X	X
	F4.5 Document the evaluation's costs, including time, human resources, expenditures, infrastructure support, and foregone opportunities		X	
	F4.6 Document the evaluation's benefits, including contributions to program improvement, future funding, better informed stakeholders, and dissemination of effective services	X		X
F4 – Resource Use - Total		**3**	**4**	**4**
Propriety				
P1. Responsive and Inclusive Orientation	P1.1 Acquire and take account of knowledge of the program environment's history, significant events, culture, and other factors affecting the program and its evaluation	X	X	X
	P1.2 Identify stakeholders broadly, gather useful information from them, and include them, as appropriate, in decisions about the evaluation's purposes, questions, and design	X	X	X

Standard		Stake Holders	Meta Evaluator	Evaluator
	P1.3 Engage and serve the full range of stakeholders in an even-handed manner, regardless of their politics, personal characteristics, status, or power	X	X	X
	P1.4 Design and schedule the evaluation to provide multiple opportunities for stakeholders to be involved, contribute, and be heard throughout the evaluation process	X	X	X
	P1.5 Be open to and thoughtfully consider stakeholders' contradictory views, interests, and beliefs regarding the program's prior history, goals, status, achievements, and significance	X	X	X
	P1.6 Avert or counteract moves by powerful stakeholders to dominate in determining evaluation purposes, questions, and procedures and interpreting outcomes	X		
P1 – Responsive and Inclusive Orientation - Total		**6**	**5**	**5**
P2. Formal Agreements	P2.1 Negotiate evaluation-related obligations, with the client, including what is to be done, how, by whom, when, and at what cost	NA	X	X
	P2.2 Make ethical, legal, and professional stipulations and obligations explicit and binding regarding such evaluation matters as evaluation purposes and questions, confidentiality/anonymity of data, editorial authority, release of reports, evaluation follow-up activities, cooperation of program staff, funds and in-kind resources, and provision for a metaevaluation	NA		X
	P2.3 Employ the contract negotiation process to strengthen trust in communications through stakeholder consultation and, unless restricted by laws or regulations, allowing stakeholders to review the printed agreement	NA		X
	P2.4 Ensure that formal evaluation agreements conform to federal, tribal, state, or local requirements, statutes, and regulations	NA		
	P2.5 Employ negotiated agreements to monitor, track, and assure effective implementation of specific duties and responsibilities	NA		X
	P2.6 Revisit evaluation agreements over time and negotiate revisions as appropriate	NA	X	
P2 – Formal Agreements - Total		**NA**	**2**	**4**
P3. Human Rights and Respect	P3.1 Adhere to applicable federal, state, local, and tribal regulations and requirements, including those of Institutional Review Boards, local/tribal constituencies, and ethics committees that authorize consent for conduct of research and evaluation studies		X	X
	P3.2 Take the initiative to learn, understand, and respect stakeholders' cultural and social	X	X	X

Standard		Stake Hold ers	Meta Eval uator	Eval uator
	backgrounds, local mores, and institutional protocols			
	P3.3 Make clear to the client and stakeholders the evaluator's ethical principles and codes of professional conduct, including the standards of the Joint Committee on Standards for Educational Evaluation	X	X	X
	P3.4 Institute and observe rules, protocols, and procedures to ensure that all evaluation team members will develop rapport with and consistently manifest respect for stakeholders and protect their rights	X	X	X
	P3.5 Make stakeholders aware of their rights to participate, withdraw, or challenge decisions that are being made at any time during the evaluation process	X	X	X
	P3.6 Monitor the interactions of evaluation team members and stakeholders and act as appropriate to ensure continuing, functional, and respectful communication and interpersonal contacts throughout the evaluation	X	X	X
P3 – Human Rights and Respect - Total		**5**	**6**	**6**
P4. Clarity and Fairness	P4.1 Develop and communicate rules that assure fairness and transparency in deciding how best to allocate available evaluation resources to address the possible competing needs of different evaluation stakeholders	X		
	P4.2 Assure that the evaluation's purposes, questions, procedures, and findings are transparent and accessible by all right-to-know audiences	X	X	X
	P4.3 Communicate to all stakeholders the evaluation's purposes, questions, and procedures and their underlying rationale	X	X	X
	P4.4 Make clear and justify any differential valuing of any stakeholders' evaluation needs over those of others		X	X
	P4.5 Carefully monitor and communicate to all right-to-know audiences the evaluation's progress and findings and do so throughout all phases of the evaluation	X	X	X
	P4.6 Scrupulously avoid and prevent any evaluation-related action that is unfair to anyone	X		X
P4 – Clarity and Fairness - Total		**5**	**4**	**5**
P5.	P5.1 Identify and disclose to all stakeholders the legal and contractual constraints under	X	X	X

Standard		Stake Hold ers	Meta Eval uator	Eval uator
Trans-parency and Disclosure	which the evaluation's information can be released and disseminated			
	P5.2 Maintain open lines of communication with and be accessible to, at least representatives of, the full range of stakeholders throughout the evaluation, so they can obtain the information which they are authorized to review	X	X	X
	P5.3 Before releasing the evaluation's findings, inform each intended recipient of the evaluation's policies— regarding such matters as right-to-know audiences, human rights, confidentiality, and privacy—and, as appropriate, acquire her or his written agreement to comply with these policies	X	X	X
	P5.4 Provide all stakeholders access to a full description and assessment of the program, e.g., its targeted and actual beneficiaries; its aims, structure, staff, process, and costs; and its strengths, weaknesses, and side effects	X	X	X
	P5.5 Provide all stakeholders with information on the evaluation's conclusions and limitations	X	X	X
	P5.6 Provide all right-to-know audiences with access to information on the evaluation's sources of monitory and in-kind support		X	X
P5 – Transparency and Disclosure - Total		5	6	6
P6. Conflicts of Interest	P6.1 Throughout the evaluation process search for potential, suspected, or actual conflicts of interest	NA	X	X
	P6.2 Search for conflicts involving a wide range of persons and groups, e.g., those associated with the client, the program's financial sponsor, program recipients, area residents, the evaluator, and other stakeholders	NA		X
	P6.3 Search for various kinds of conflicting interests, including prospects for financial gains or losses, competing program goals, alternative program procedures, alternative evaluation approaches, and alternative bases for interpreting findings	NA		
	P6.4 Take appropriate steps to manage identified conflicts so that the evaluation maintains integrity and high quality	NA		X
	P6.5 Attend to conflicts of interest through effective communication with the client and other pertinent parties and in a spirit of mutual and deliberate understanding and learning	X		X
	P6.6 Document and report identified conflicts of interest, how they were addressed, and	X	X	X

Standard		Stake Hold ers	Meta Eval uator	Eval uator
	how they affected the evaluation's soundness			
P6 – Conflicts of Interest		**NA**	**2**	**5**
P7. Fiscal Responsi- bility	P7.1 Plan and obtain approval of the evaluation budget before beginning evaluation implementation	NA		X
	P7.2 Be frugal in expending evaluation resources	NA		X
	P7.3 Employ professionally accepted accounting and auditing practices	NA		
	P7.4 Maintain accurate and clear fiscal records detailing exact expenditures, including adequate personnel records concerning job allocations and time spent on the job	NA		
	P7.5 Make accounting records and audit reports available for oversight purposes and inspection by stakeholders	NA		
	P7.6 Plan for and obtain appropriate approval for needed budgetary modifications over time or because of unexpected problems	NA		
P7 – Fiscal Responsibility - Total		**NA**	**0**	**2**
Accuracy				
A1. Justified Conclusions and Decisions	A1.1 Address each contracted evaluation question based on information that is sufficiently broad, deep, reliable, contextually relevant, culturally sensitive, and valid	X	X	X
	A1.2 Derive defensible conclusions that respond to the evaluation's stated purposes, e.g., to identify and assess the program's strengths and weaknesses, main effects and side effects, and worth and merit	X	X	X
	A1.3 Limit conclusions to the applicable time periods, contexts, purposes, and activities	X	X	X
	A1.4 Identify the persons who determined the evaluation's conclusions, e.g., the evaluator using the obtained information plus inputs from a broad range of stakeholders	X	X	X
	A1.5 Identify and report all important assumptions, the interpretive frameworks and values employed to derive the conclusions, and any appropriate caveats	X	X	X
	A1.6 Report plausible alternative explanations of the findings and explain why rival explanations were rejected	X	X	
A1 – Justified Conclusions and Decisions - Total		**6**	**6**	**5**
A2.	A2.1 Through communication with the full range of stakeholders develop a coherent,	X	X	X

Standard		Stake Holders	Meta Eval uator	Eval uator
Valid Information	widely understood set of concepts and terms needed to assess and judge the program within its cultural context			
	A2.2 Assure—through such means as systematic protocols, training, and calibration--that data collectors competently obtain the needed data	NA	X	X
	A2.3 Document the methodological steps taken to protect validity during data selection, collection, storage, and analysis	NA	X	
	A2.4 Involve clients, sponsors, and other stakeholders sufficiently to ensure that the scope and depth of interpretations are aligned with their needs and widely understood	X	X	X
	A2.5 Investigate and report threats to validity, e.g., by examining and reporting on the merits of alternative explanations	NA	X	
	A2.6 Assess and report the comprehensiveness, quality, and clarity of the information provided by the procedures as a set in relation to the information needed to address the evaluation's purposes and questions	NA	X	X
A2 – Valid Information - Total		NA	**6**	**4**
A3. Reliable Information	A3.1 Determine, justify, and report the needed types of reliability—e/g., test-retest, findings from parallel groups, or ratings by multiple observers—and the acceptable levels of reliability	NA		
	A3.2 In the process of examining, strengthening, and reporting reliability, account for situations where assessments are or may be differentially reliable due to varying characteristics of persons and groups in the evaluation's context	NA		X
	A3.3 Assure that the evaluation team includes or has access to expertise needed to investigate the applicable types of reliability	NA	X	
	A3.4 Describe the procedures used to achieve consistency	NA	X	X
	A3.5 Provide appropriate reliability estimates for key information summaries, including descriptions of programs, program components, contexts, and outcomes	NA		X
	A3.6 Examine and discuss the consistency of scoring, categorization, and coding and between different sets of information, e.g., assessments by different observers	NA	X	
A3 – Reliable Information - Total		NA	**3**	**3**
A4.	A4.1 Describe all important aspects of the program—e.g., goals, design, intended and actual	X	X	X

Standard		Stake Holders	Meta Evaluator	Evaluator
Explicit Program and Context Descriptions	recipients, components and subcomponents, staff and resources, procedures, and activities—and how these evolved over time			
	A4.2 Describe how people in the program's general area experienced and perceived the program's existence, importance, and quality	X	X	X
	A4.3 Identify any model or theory that program staff invoked to structure and carry out the program	X	X	X
	A4.4 Define, analyze, and characterize contextual influences that appeared to significantly influence the program and that might be of interest to potential adopters, including the context's technical, social, political, organizational, and economic features	X	X	X
	A4.5 Identify any other programs, projects, or factors in the context that may affect the evaluated program's operations and accomplishments	X	X	X
	A4.6 As appropriate, report how the program's context is similar to or different from contexts where the program is expected to or reasonably might be adopted	X	X	X
A4 – Explicit Program and Context Descriptions - Total		**6**	**6**	**6**
A5. Information Manage- ment	A5.1 Select information sources and procedures that are most likely to meet the evaluation's needs for accuracy and be respected by the evaluation's client group	NA	X	X
	A5.2 Ensure that the collection of information is systematic, replicable, adequately free of mistakes, and well documented	NA	X	X
	A5.3 Establish and implement protocols for quality control of the collection, validation, storage, and retrieval of evaluation information	NA	X	X
	A5.4 Document and maintain both the original and processed versions of obtained information	NA	X	X
	A5.5 Retain the original and analyzed forms of information as long as authorized users need it	NA		X
	A5.6 Store the evaluative information in ways that prevent direct and indirect alterations, distortions, destruction, or decay	NA		X
A5 – Information Management - Total		**NA**	**4**	**6**
A6. Sound	A6.1 Create or select a logical framework that provides a sound basis for studying the subject program, answering the evaluation's questions, and judging the program and its	X	X	X

Standard		Stake Holders	Meta Evaluator	Evaluator
Designs and Analyses	components			
	A6.2 Plan to access pertinent information sources and to collect a sufficient breadth and depth of relevant, high quality quantitative and qualitative information in order to answer the evaluation's questions and judge the program's value	X	X	X
	A6.3 Delineate the many specific details required to collect, analyze, and report the needed information	NA	X	X
	A6.4 Develop specific plans for analyzing obtained information, including clarifying needed assumptions, checking and correcting data and information, aggregating data, and checking for statistical significance of observed changes or differences in program recipients' performance	NA	X	X
	A6.5 Buttress the conceptual framework and technical evaluation design with concrete plans for staffing, funding, scheduling, documenting, and metaevaluating the evaluation work	NA	X	X
	A6.6 Plan specific procedures to avert and check for threats to reaching defensible conclusions, including analysis of factors of contextual complexity, examination of the sufficiency and validity of obtained information, checking on the plausibility of assumptions underlying the evaluation design, and assessment of the plausibility of alternative interpretations and conclusions	NA	X	X
A6 – Sound Designs and Analyses - Total		**NA**	**6**	**6**
A7. Explicit Evaluation Reasoning	A7.1 Clearly describe all the assumptions, criteria, and evidence that provided the basis for judgments and conclusions	X	X	X
	A7.2 In making reasoning explicit, begin with the most important questions, then, as feasible, address all other key questions, e.g., those related to description, improvement, causal attributions, accountability, and costs related to effectiveness or benefits	X	X	X
	A7.3 Document the evaluation's chain of reasoning, including the values invoked so that stakeholders who might embrace different values can assess the evaluation's judgments and conclusions	X	X	X

212

Standard		Stake Holders	Meta Evaluator	Evaluator
	A7.4 Examine and report how the evaluation's judgments and conclusions are or are not consistent with the possibly varying value orientations and positions of different stakeholders	X		X
	A7.5 Identify, evaluate, and report the relative defensibility of alternative conclusions that might have been reached based on the obtained evidence	X		
	A7.6 Assess and acknowledge limitations of the reasoning that led to the evaluation's judgments and conclusions	X	X	X
A7 – Explicit Evaluation Reasoning		**6**	**4**	**5**
A8. Communicating and Reporting	A8.1 Reach a formal agreement that the evaluator will retain editorial authority over reports	NA		
	A8.2 Reach a formal agreement defining right-to-know audiences and guaranteeing appropriate levels of openness and transparency in releasing and disseminating evaluation findings	NA		X
	A8.3 Schedule formal and informal reporting in consideration of user needs, including follow-up assistance for applying findings	X		X
	A8.4 Employ multiple reporting mechanisms, e.g., slides, dramatizations, photographs, PowerPoint©, focus groups, printed reports, oral presentations, telephone conversations, and memos	X	X	X
	A8.5 Provide safeguards, such as stakeholder reviews of draft reports and translations into language of users, to assure that formal evaluation reports are correct, relevant, and understood by representatives of all segments of the evaluation's audience	X		X
	A8.6 Consistently check and correct draft reports to assure they are impartial, objective, free from bias, responsive to contracted evaluation questions, accurate, free of ambiguity, understood by key stakeholders, and edited for clarity	X	X	X
A8 – Explicit Evaluation Reasoning		**NA**	**2**	**5**
Evaluation Accountability				
E1. Evaluation Docu Ment	E1.1 Document and preserve for inspection the following: Contract or memorandum of agreement that governed the evaluation	NA		
	E1.2 Evaluation plan, including evaluation tools and resumes of key evaluation staff	NA	X	X
	E1.3 Evaluation budget and cost records	NA		

Standard		Stake Holders	Meta Evaluator	Evaluator
ation	E1.4 Reports, including interim and final reports, the evaluation's internal metaevaluation report, and, if obtained, a copy of the external metaevaluation report	NA	X	X
	E1.5 Other information determined to be needed by reviewers, such as technical data on the employed evaluation tools, a glossary of pertinent theoretical and operational definitions involved in the evaluation, a description of the subject program, a record of stakeholder involvement, and news accounts related to the evaluation	NA	X	X
	E1.6 Evidence of the evaluation's consequences, including stakeholders' uses of findings	NA		X
E1 – Evaluation Documentation - Total		NA	3	4
E2. Internal Metaevaluation	E2.1 At the evaluation's beginning, determine the metaevaluation's intended users and uses (e.g., formative and summative)	NA		X
	E2.2 Develop a plan for obtaining, processing, and reporting a sufficient scope and depth of information to assess the evaluation's utility, feasibility, propriety, and accuracy and address the intended users' needs for timely metaevaluation feedback and reports	NA	X	X
	E2.3 Assign responsibility for documenting and assessing the evaluation's plans, process, findings, and impacts and budget sufficient resources to carry out the internal metaevaluation	NA		X
	E2.4 Maintain and make available for inspection a record of all internal metaevaluation steps, information, analyses, costs, and observed uses of the metaevaluation findings	NA		X
	E2.5 Reach, justify, and report Judgments of the evaluation's adherence to all of the metaevaluation	NA	X	X
	E2.6 Make the internal metaevaluation findings available to all authorized users	NA		X
E2 – Internal Metaevaluation - Total		NA	2	6
E3. External Meta-evaluation	E3.1 Confirm through exchange with key stakeholders the need for an external assessment of the evaluation and the purposes it should serve (e.g., formative or summative	NA	X	
	E3.2 Stipulate that these and possibly additional standards will be used to assess and judge the evaluation	NA	X	X
	E3.3 Select, recruit, and reach a formal agreement with an external metaevaluator who possesses an independent perspective, appropriate expertise, and freedom from possibly compromising connections or interests	NA	X	X

Standard		Stake Holders	Meta Evaluator	Evaluator
	E3.4 Assure that the external metaevaluation is adequately planned, staffed, and funded	NA	X	X
	E3.5 Provide the external metaevaluator with access to information and personnel required to conduct a thorough, defensible metaevaluation that serves the intended purposes	NA	X	X
	E3.6 Assure that the metaevaluation will be subjected to appropriate quality control and that the metaevaluator will deliver as part of the metaevaluation report an attestation of its adherence to the metaevaluation standards	NA	X	X
E3 – External Metaevaluation - Total		**NA**	**6**	**5**

REFERENCES

Abell Foundation, The (2008). One-to-One computing in public schools: Lessons from "laptops for all" programs. Retrieved from http://abell.org/pubsitems/ed_onetoone_908.pfd

Alkin, M.C. (1991). Evaluation theory development. In M.W. McLaughlin & D.C. Phillips (Eds.) *Evaluation and education: At quarter-century* (90th yearbook of the National Society for the Study of Education, Part II), p. 91-112. Chicago: University of Chicago Press.

Alkin, M.C. & Christie, C.A. (2004). An evaluation theory tree. In M.C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences,* p. 12-65. Thousand Oaks, CA: Sage.

American Evaluation Association. (2012). Downloaded from http://www.eval.org/evaluationdocuments/progeval.html.

Ashworth, A., Cebulla, A., Greenberg, D., & Walker, A. (2004). Metaevaluation: Discovering what works best in welfare provision. *Evaluation,* 10, 193-216.

Blumenthal, P.G. (2003). Evaluating ubiquitous access to technology: Teacher and administrator perceptions of effective classroom practice. (Doctoral dissertation). Retrieved from http://proquest.umi.com/pqdweb?index=97&did=765247591&SrchMode=1&sid=10&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1328198968&clientId=14288

Brandon, P. (1998). A metaevaluation of schools' methods for selecting site-managed projects. *Studies in Educational Evaluation,* 24, 213-228.

Brethower, D. (2012). Foreword – Fundamentals of performance technology. In D.M. Van Tiem, J.L. Moseley & J.C. Dessinger *Fundamentals of performance improvement: Optimizing results through people, process and organizations, 3rd ed.* San Francisco: Pfeiffer.

Boruch, R., McSweeney, A., & Soderstrom, E. (1978). Randomized field experiments for program planning, development and evaluation. *Evaluation Quarterly,* 2(4), 655-695.

Brinkerhoff, R.O. (2003). *The success case method: Find out quickly what's working and what's not.* San Francisco: Berrett-Koehler.

Brinkerhoff, R.O., Brethower, D.M., Hluchyj, T., & Nowakowski, J.R. (1983). *Program evaluation: A practitioner's guide for trainers and educators.* Boston: Kluwer-Nijhoff.

Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Chelimsky, E. (1995). The political environment of evaluation and what it means for the development of the field, *Evaluation Practice,* 16(3), 215-225.

Chandrasekhar, V.S. (2009). Promoting 21st century learning: A case study of the changing role of teachers in one-to-one laptop classrooms. (Doctoral dissertation). Downloaded from ProQuest Dissertations. (AAT 3364982)

Chen, J., & Rossi, P. (1983). Evaluating with sense: The theory-driven approach, *Evaluation Review,* 7, 283-302.

Christie, C.A. & Alkin, M.C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation,* 34, p. 131-135.

Christie, C. & Alkin, M. (2012). An evaluation theory tree. In M.C. Alkin (2012) *Evaluation Roots (2nd Edition).* Thousand Oaks, CA: Sage.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Designing and analysis issues for field settings.* Chicago: Rand McNally.

Cooksy, L.J. (1999). The metaevaluand: The evaluation of project TEAMS. *American Journal of Evaluation,* 20(1), 123-136.

Cooksy, L.J. & Caracelli, V.J. (2009). Metaevaluation in practice: Selection and application of criteria. *Journal of multidisciplinary evaluation,* 6(11), 1-15.

Cousins, J., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80,* 5-23.

Cronbach, L. & Associates (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements.* San Francisco: Jossey-Bass.

Crotty, M. (1998). *The foundations of social research.* London: Sage.

Dalgarno, N.J. (2009). Compulsory laptop programs: Teachers' responses to the adoption and implementation process. (Doctoral dissertation). Retrieved from ProQuest Dissertations. (AAT NR65424)

Dessinger, J.C. & Moseley, J.L. (2004). *Confirmative evaluation: Practical strategies for valuing continuous improvement.* San Francisco: Pfeiffer.

Ebel, R.L. (1967). Some limitations of basic research in education. *Phi Delta Kappan,* 49, 81-84.

Eisner, E. (1976). Educational connoisseurship and criticism: Their form and function in educational evaluation. *Journal of Aesthetic Evaluation or Education,* 10, 135-150.

Eisner, E. (1979). *The educational imagination: On the design and evaluation of school programs.* New York: MacMillan.

Fetterman, D. (1994). Steps of empowerment education: From California to Cape Town. *Evaluation and Program Planning,* 17(3), 305-313.

Finn, J. (1960). Professionalizing the audio-visual field, *Educational Technology Research and Development,* 1, 1, 6-17.

Fitzpatrick, J.L., Sanders, J.R., Worthen, B.R. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines, 3rd ed.* Boston: Pearson Education.

Fitzpatrick, J.L., Sanders, J.R., Worthen, B.R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines, 4ᵗʰ ed.* Boston: Pearson Education.

Fleiss, J. L. (1971) Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76 (5), 378–382.

Furr, R.M., & Bacharach, V.R. (2008). *Psychometrics: An introduction.* Los Angeles: Sage.

Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., Vermeersch, C.M. (2010). *Impact evaluation in practice.* New York: World Bank.

Gilbert, T.F. (1978). *Human competence: Engineering worthy performance.* New York: McGraw Hill.

Grasso, P. (1999). Metaevaluation of an evaluation of reader focused writing the for veterans benefits administration. *American Journal of Evaluation.* 20, 355-370.

Greene, J.C. (2005). A value-engaged approach for evaluating the Bunche-Da Vinci learning academy, *New Directions for Evaluation,* 106, 27-47.

Guba, E.G., & Lincoln, Y.S. (1981). *Effective Evaluation.* San Francisco: Jossey-Bass.

Guba, E.G. & Lincoln, Y.S. (2004). The roots of fourth generation evaluation: Theoretical and methodological origins, in M. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences,* Thousand Oaks, CA: Sage.

Guerra-Lopez, I. (2007a). *Evaluating impact: Evaluation and continual improvement for performance improvement practitioners.* Champaign, IL: Human Resource Development Press.

Guerra-Lopez, I. (2007b). Evaluating impact: Building a case for demonstrating the worth of performance improvement interventions. *Performance Improvement Journal,* 46(7), 33-38.

Guerra-Lopez (2007c). Planning a responsive evaluation: Establishing solid partnerships by clarifying expectations and purpose. *Performance Improvement Journal,* 46(8).

Guerra-Lopez (2007d). A seven-step evaluation process for evaluating performance improvement interventions. *Performance Improvement Express,* no. 10. Retrieved from http://www.performancexpress.org

Guerra-Lopez, I. (2008). *Performance evaluation: Proven approaches for improving program and organizational performance.* San Francisco: Jossey-Bass.

Guerra-Lopez, I. (2010). Performance tracking and management systems. In Watkins, R. & Leigh, D. (Eds). *Handbook for the selections and implementation of human performance interventions.* New York: Wiley & Sons.

Guerra-Lopez, I. (2012). The monitoring and impact evaluation process: A systematic approach to improving performance and impact, *International Journal for Environmental Science and Engineering,* 3 (30), 80-85.

Guerra-Lopez, I. & Toker, S. (2012). An application of the Impact Evaluation Process for designing a performance measurement and evaluation framework in K-12 environments. *Evaluation and Program Planning,* 35, 222-235.

Gullickson, A., Wingate, L., Lawrenz, F.A., & Coryn, C. (2006). *The NSF advanced technology education program: Final evaluation report.* Kalamazoo, MI: Western Michigan University: The Evaluation Center.

Guskey, T.R. (2003). What makes professional development effective? *Phi Delta Kappan*, 84 (10), 748-751.

Hanssen, C.E., Lawrenz, F., Dunet, D.O. (2008). Concurrent meta-evaluation: A critique, *American Journal of Evaluation,* 29, 572-582.

Harris, M.J. (2010). Impactful student learning outcomes of one-to-one student laptop programs in low socioeconomic schools. (Doctoral dissertation). Retrieved from ProQuest Dissertations. (AAT 3438093)

Henry, G.T. & Mark, M.M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation.* 97, p. 69-80.

Henry, G., Julnes, G.J., and Mark, M. (1998). *Realist evaluation.* San Francisco: Jossey-Bass.

Holcomb, L.B. (2009). Results and lessons learned from 1:1 laptop initiatives: A collective review. *Tech Trends,* 53(6), 49-55.

House, E.R. (1976). Justice in evaluation. In G.V. Glass (Ed.), *Evaluation studies review annual* (Vol 1). Beverly Hills, CA: Sage.

House, E.R. (1983). Assumptions underlying evaluation models. In G.F. Madaus, M. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation.* Boston: Kluwer-Nijhoff.

House, E.R. (1991). Evaluation and social justice: Where are we? In M.W. McLaughlin & D.C. Phillips (Eds.), *Evaluation and education: At quarter century* (90th yearbook of the National Society for the Study of Education, Part II), pp. 233-247, Chicago: University of Chicago Press.

House, E.R., & Howe, K.R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

ISPI (2012). *What is HPT?* Retrieved from http://www.ispi.org/content.aspx?id=54.

Joint Committee on Standards for Educational Evaluation. (2010). *The program evaluation standards* (3rd ed.). Thousand Oaks, CA: Sage.

Juni, S. (2007). Reliability theory. In N.J. Salkind (Ed.). *Encyclopedia of measurement and statistics, Volume 3*. pp. 834-835. Thousand Oaks, CA: Sage.

Kaufman, R.A. (2006). *Changes, choices and consequences: A guide to mega thinking and planning.* Amherst, MA: HRD Press.

Kaufman, R.A., & Clark, R. (1999). Reestablishing performance improvement as a legitimate area of inquiry, activity, and contribution: Rules of the road. *Performance Improvement, 38*(9), 13-18.

Kerr, K.A., Pane, J.F., & Barney, H. (2003). *Quaker Valley digital school district: Early effects and plans for future evaluation.* Santa Monica, CA: Rand Corporation.

King, J.A., & Stevahn, L. (2005). *Interactive evaluation in practice: Managing the interpersonal dynamics of program evaluation.* Thousand Oaks, CA: Sage.

Kirkpatrick, D.L. (1959). Techniques for evaluating training programs. *Journal of ASTD, 13*(11), 21-26.

Kirkpatrick, D.L. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Klein, J.D. (2002). Empirical research on performance improvement. *Performance Improvement Quarterly. 15*, 1, 99-110.

Landis, J.R. & Koch, G.C. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 169-174.

Levin, H.M. (2005). Cost-benefit analysis. In S. Mathison (Ed.), *Encyclopedia of Evaluation,* Thousand Oaks, CA: Sage.

Luo, H. (2011). Qualitative research on educational technology: Philosophies, methods and challenges. *International Journal of Education,* 3(2), 1-16.

Lynch, D.C., Greer, A.G., Larson, L.C., Cummings, D.M., Harriett, B.S., Dreyfus, K.S. & Clay, M.C. (2003). Descriptive metaevaluation: Case study of an interdisciplinary curriculum, *Evaluation and the health professions,* 26, 4, 447-461.

Mager, R.F. (2006). Foreword to the second edition. In J.A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, potential.* (pp. xxxvii-xliv). San Francisco: Pfeiffer.

Marrelli, A.F. (2011). A universal performance model for human performance technology: The time has come. *Performance Improvement,* 50, 9.

Mark, M.M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation,* 22(3), 457-479.

Mark, M.M., & Henry, G. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation,* 10(1), 35-37.

McLaren, Z. (2011). Effect of laptop technology on student interaction. (Doctoral dissertation). Retrieved from ProQuest Dissertations. (AAT 3451875)

Merriam, S.B. (2009). *Qualitative research: A guide to design and implementation.* San Francisco: Jossey-Bass.

Mertens, D. (1999). Inclusive evaluation: Implications of transformative theory for evaluation, *American Journal of Evaluation,* 20, 1, 1-14.

Mertens, D. (2001). Inclusivity and transformation: Evaluation in 2010, *American Journal of Evaluation,* 22, 367-374.

Miller, R.L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation,* 31(3), 390-399.

Mills, C.E. (2010). Laptop education: Impact on learning and achievement. (Doctoral dissertation). Retrieved from ProQuest Dissertations. (AAT 3445660)

Newman, D. (2011). An empirical validation of Guskey's Professional Development Evaluation Model using six years of student and teacher level reading data. Ph.D. dissertation, Cleveland State University, United States -- Ohio. Retrieved January 23, 2012, from Dissertations & Theses: Full Text. (Publication No. AAT 3446485).

Nilsson, N., & Hogben, D. (1983). Metaevaluation. In E.R. House (Ed.), *Philosophy of evaluation.* New Directions for Program Evaluation, 19, 83-97. San Francisco: Jossey-Bass.

Owens, T.R. (1973). Educational evaluation by adversary proceeding. In E.R. House (Ed.). *School evaluation: The politics and process.* Berkeley, CA: McCutchan.

Partnership for 21st Century Skills. (2012) Retrieved from http://www.p21.org

Patton, M.Q. (1986). Utilization-focused evaluation (2nd ed). Beverly Hills, CA: Sage.

Patton, M.Q. (2008). *Utilization-focused evaluation (*4th ed.). Thousand Oaks, CA: Sage.

Pershing, J. (2006). *Handbook of human performance technology: Principles, practices, potential.* San Francisco: Jossey-Bass.

Pershing, J., Lee, J., & Cheng, J. (2008). Current status, future trends, and issues in human performance technology, part 2: Models, influential disciplines, and research and development, *Performance Improvement,* 47, 2, 7-15.

Phillips, J. (1997). *Return on investment in training and performance improvement programs.* Burlington, MA: Butterworth-Heinemann.

Pogany, W.S. (2009). Evolving pedagological perceptions of teachers integrating ubiquitous computing in their classrooms: A case study of the South Dakota laptop program. (Doctoral dissertation). Downloaded from ProQuest Dissertations. (AAT 3388228)

Preskill, H., & Torres, R.T. (Eds.). (2001). The learning dimension of evaluation use, *New Directions for Evaluation (Vol. 88).* San Francisco: Jossey-Bass.

Provus, M. (1971). *Discrepancy evaluation.* Berkeley, CA: McCutcheon.

Queener, R.C. (2011). Effects of laptop initiatives on student achievement in mathematics. (Doctoral Dissertation). Retrieved from ProQuest Dissertations. (AAT 3457213)

Reiser, R.A., Dempsey, J.V. (2012). *Trends and issues in instructional design and technology, (3*rd* ed).* Upper Saddle River, NJ: Pearson.

Rockman, S. (1998). Powerful tools for schooling: Second year study of the laptop program. San Francisco: CA.

Ross, S.M., Morrison, G.R., & Lowther, D.L. (2010). Educational technology research past and present: Balancing rigor and relevance to impact school learning. *Contemporary Educational Technology, 1,* 17-35.

Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach (6*th* ed.).* Thousand Oaks, CA: Sage.

Ryan, K.E., & Schwandt, T.A. (Eds.). (2002). *Exploring evaluator role and identity.* Greenwich, CT: Information Age Publishing.

Savnye, W.C. & Robinson, R.S. (2004). Qualitative research issues and methods: An introduction for educational technologists. In Jonassen, D.H. (2004). *Handbook of Research on Educational Communications and Technology,* 1045-1072. Mahwah, NJ: Erlbaum.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.). *Curriculum evaluation.* American Educational Research Association Monograph Series on Evaluation, No. 1, 39-83. Chicago: Rand McNally.

Scriven, M. (1972). Prose and cons about goal-free evaluation. *Evaluation Comment: The Journal of Educational Evaluation,* 3(4), 1-7. UCLA, Center for the Study of Evaluation.

Scriven, M. (1991). *Evaluation thesaurus (4th ed.).* Newbury Park, CA: Sage.

Scriven, M. (1993). *Hard-won lessons in program evaluation*. New Directions for Program Evaluation*,* 58, 1-107. San Francisco: Jossey-Bass.

Skevakis, A. (2010). An analysis of teacher perceptions of a principal's leadership behaviors associated with the integration of a one-to-one laptop program for students in a parochial secondary school. (Doctoral dissertation). Retrieved from ProQuest Dissertations (AAT 3404212)

Smith, N.L. (2007). Empowerment evaluation as ideology. *American Journal of Evaluation,* 28, 169-178.

Staib, M. (2011). An examination of 21st century tools in student work in a grade 6 classroom. (Doctoral dissertation). Retrieved from ProQuest Dissertations (AAT 3477342)

Stake, R. (1975). *Program evaluation, particularly responsive evaluation* (Occasional Paper No. 5). Kalamazoo: Western Michigan University Evaluation Center.

Stake, R. (1981). Case study methodology: An epistemological advocacy. In W.W. Welsh (Ed.), *Complementary methods for research in education.* Washington, DC: American Educational Research Association.

Stolovitch, H.D. (2000). Human performance technology: Research and theory to practice. *Performance Improvement,* 39(4), 7-16.

Stufflebeam, D.L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education,* 5, 19-25.

Stufflebeam, D.L. (1983). The CIPP model for program evaluation. In G.F. Madaus, M.S. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on education and human services evaluation,* 117-141. Boston: Kluwer-Nijhoff.

Stufflebeam, D.L. (2000). The methodology of metaevaluation as reflected in metaevaluations by the Western Michigan University Evaluation Center, *Journal of Personnel Evaluation in Education,* 14(1), 95-125.

Stufflebeam, D.L. (2001a). Evaluation models. *New Directions for Evaluation,* 89, 7-98.

Stufflebeam, D.L. (2001b). The metaevaluation imperative. *American Journal of Evaluation,* 22, 2, 183-209.

Stufflebeam, D.L. (2011). *Program evaluations metaevaluation checklist.* Retrieved from: http://www.wmich.edu/evalctr/checklists/metaevaluation/

Stufflebeam, D.L. & Coryn, C.L.S. (2012). *Evaluation theory, models, and applications (2nd ed.).* San Francisco, CA: Jossey-Bass.

Sugrue, B., & Stolovitch, H. (2000). Appropriate inquiry in performance technology. *Performance Improvement,* 39(1), 33-36.

Tyler, R.W. (1942). General statement on evaluation. *Journal of Educational Research,* 35, 492-501.

Van Tiem, D.M., Moseley, J.L. & Dessinger, J.C. (2012). *Fundamentals of performance improvement: Optimizing results through people, process and organizations, 3rd ed.* San Francisco: Pfeiffer.

Weiss, C. (1991). Evaluation research in the political context: Sixteen years and four administrations later. In M.W. McLaughlin & D.C. Phillips (Eds.). *Evaluation and education: At quarter century (90*^{*th*} *yearbook of the National Society for the Study of Education, Part II).* 211-231. Chicago: University of Chicago Press.

Wholey, J.S. (1983). *Evaluation and effective public management.* Boston: Little, Brown.

Wingate, L. (2009). The program evaluation standards applied for metaevaluation purposes: Investigating interrater reliability and implications for use. (Doctoral dissertation, Western Michigan University). Retrieved from http://search.proquest.com.proxy.lib.wayne.edu/docview/305029057/fulltextPDF/139F87 A0A096A856D74/2?accountid=14925

Wolf, R.L. (1979). The use of judicial evaluation methods in the formulation of educational policy. *Educational Evaluation and Policy Analysis,* 1(93), 19-28.

Woodside, A., & Sakai, M. (2001). Metaevaluation of performance audits of government tourism marketing programs. *Journal of Travel Research,* 39, 369-379.

Wong, S.E. (2010). Single-case evaluation designs for practitioners. *Journal of Social Service Research,* 36, 248-259.

Yarbrough, D.B., Shulha, L.M., Hopson, R.K., and Caruthers, F.A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* 3rd ed.). Thousand Oaks, CA: Sage.

Zucker, A. (2004). Developing a research agenda for ubiquitous computing in schools. *Journal of Educational Computing Research,* 30(4), 371-386.

ABSTRACT

AN EMPIRICAL EXAMINATION AND METAEVALUATION
OF THE IMPACT EVALUATION PROCESS

by

ANNE MAECHLING BLAKE

May 2013

Advisor: Dr. James L. Moseley

Major: Instructional Technology

Degree: Doctor of Philosophy

This dissertation empirically examines the Guerra-Lopez (2007a) Impact Evaluation Process (IEP), which is a prescriptive program evaluation model. Since there is no generally accepted process for arriving at final judgments about the usefulness, appropriateness, effectiveness, reliability, and validity of evaluation models, this study used a combination of approaches to begin to build a body of evidence about the effectiveness of the IEP. Primarily, the study used Stufflebeam's (2011) recently revised *Program Evaluations Metaevaluation Checklist* to examine the model. The *Checklist* is based on the Joint Committee on Standards for Educational Evaluation's (2010) *Program Evaluation Standards*. Fitzpatrick, Sanders, *&* Worthen (2011) recommend selecting a subset of these standards to use when evaluation a design. Additionally, the study used Miller's (2010) framework for empirically evaluating how evaluation theory informs practice. First, through a study of three evaluation theory classification schemes, the researcher identifies where the IEP fits among other common evaluation models. Next, in order to reach a judgment based on the model's application in the real world, the researcher conducted an impact evaluation on a 1:1 technology program at a secondary school

using the model. The process used to conduct the evaluation is discussed in detail. As part of the process, the researcher developed an operationalized version of the model. Based on these standards and Stufflebeam's (2011) *Checklist* scoring method, the evaluator and a professional metaevaluator rated the Impact Evaluation Process as "Very Good".

AUTOBIOGRAPHICAL STATEMENT

Anne Maechling Blake received a degree in Economics and Computer Science from Georgetown University. She graduated from General Electric's Financial Management Program and worked in accounting and operations at General Electric and The Stroh Brewery Company. Prior to returning to school full-time to pursue a master's degree, Anne was responsible for the purchase and on-time delivery of all glass and paper products (a.k.a. beer bottles and labels) for Stroh's seven breweries nationwide. Anne earned a master's degree in Higher and Adult Continuing Education at the University of Michigan, and returned briefly to Stroh's as a member of the Human Resources department, until deciding to become a stay-at-home parent. During the next fifteen years, Anne participated in a wide variety of volunteer activities that focused on serving children and families living in poverty. After returning to school to pursue an Education Specialist Certificate and a doctorate, Anne worked for several years as a Graduate Research Assistant in Instructional Technology and most recently, as a part-time faculty member at Wayne State University.