Structural Equation Modeling as a Tool for Multisite Evaluation

Carl Edward Hanssen
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations
Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1104
STRUCTURAL EQUATION MODELING AS A TOOL FOR
MULTISITE EVALUATION

by

Carl Edward Hanssen

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Educational Studies

Western Michigan University
Kalamazoo, Michigan
June 2004
Two fundamental purposes exist for program evaluation: to document program results and to improve programs. These purposes are commonly called summative and formative evaluation. The fundamental questions related to these purposes are (1) what occurred in the program? and (2) how can the program be improved? The answer to the second question implies the need to explain why program results occurred.

This dissertation developed an approach to support formative evaluation and answering the question of why program results occur. This approach integrated multisite evaluation, theory-based evaluation and structural equation modeling. The context for this dissertation was the National Science Foundation's Advanced Technological Education (ATE) program. ATE program logic indicated that project characteristics and organizational practices were positively related to levels of collaboration; levels of collaboration were positively related to productivity in materials development, professional development, and program improvement; and program improvement was positively related to student impact.

This study questioned (1) if the ATE program logic model fit the empirical data available from an annual survey of ATE projects, (2) if the model could be fitted, what could be concluded about the relationships between program characteristics
and results, and (3) if the model fit was not optimal, could the program logic model be modified to improve the fit.

Robust maximum likelihood estimation, as implemented in LISREL 8.54, was used to determine the model fit. The ATE program logic model provided an overall acceptable fit to the data, though standardized path coefficients indicated that some components of the model were not supported. Results also suggested that the measurement of program characteristics, organizational practices, and materials development was poor. An alternative model with collaboration driving program improvement and professional development results, and program improvement driving student impact provided the strongest fit to the data.

The implications of this study were that structural equation modeling represents a promising analytical approach to support formative evaluation in multisite evaluation. Challenges in implementing the approach are articulating and measuring the program logic model, and achieving sufficient sample size. Recommendations for evaluation practitioners and future research are provided.
ACKNOWLEDGMENTS

It has taken me a bit longer than normal to complete my education. I began my doctoral studies in 1994 about eight years after I first earned an undergraduate degree from Northwestern and were it not for a six-plus year side trip to Andersen, I might have earned the degree before the turn of the century. As it was, I did what I did and am here now, with few regrets and a whole lot of experience which I believe has helped me tremendously. My pursuit of a doctorate has sometimes been stimulating, boring, frustrating, and exciting. The constant along the way, however, has been the opportunity to meet and interact with interesting people. I want to thank them all now as I am very much a reflection of their influence on me.

First, my dissertation committee, most notably Dr. Brooks Applegate, deserves special thanks. Dr. Applegate ushered me through this process, challenged me, and has helped me appreciate some of the unresolved tensions inherent in evaluation. Dr. Frances Lawrenz has been supportive of me along the way, encouraged my work, and has been remarkable for the speed and clarity with which she provides feedback. Dr. Jianping Shen has provided me with support and thoughtful recommendations at important points in time.

Second, my colleagues at The Evaluation Center at WMU deserve thanks. Dr. Arlen Gullickson brought me back to the Center in 2003 and offered the avenue for me to complete my studies through the evaluation of ATE. Dr. Daniel Stufflebeam was my first mentor in evaluation. Though I abandoned him back in 1996, he welcomed me back in 2003, encouraged me to finish up, and gave me
invaluable advice. Thank you Dan! Jane Davidson, Chris Nelson (thanks Chris for the SEM book!), and Dr. James Sanders all encouraged me. Sally Veeder and Mary Ramlow have been great friends. Chris Coryn has been an invaluable research assistant who has suffered through most of my thinking aloud. Chris has a bright future and I look forward to working with him.

Third, my colleagues at Arthur Andersen deserve special mention. Of course, the demise of that organization in 2002 was a catalyst in my completing the degree, but I believe that a stronger group of evaluators has rarely been assembled in one place in such numbers. It is in many ways sad that the organization fell in such a public way as that has diminished much of what we accomplished. All along the way, they kidded me about my lack of education, but they were also respectful of my contributions, preferring to judge me on merit rather than credentials. I will forget some, but they are Oliver Cummings, John Wallace, John Mattox, John Richter, Darryl Jinkerson, Jennifer Dewey, Nancy DeLay, Judy Powills, Rose Setze, Dave Gibson, Constance Filling, Ellen Johnson, Lisa Lund, Mary Bicek, and Coley O'Brien. Special thanks to Oliver who has been a true mentor and friend.

My family has earned the final slot in these acknowledgements as they have perhaps suffered the most over the past year. My wife Amy has been unflappable in her support of me over 10 years of marriage (about the time it has taken me to complete this degree!). She has regularly set her own needs aside for mine and I am sure that I can never repay that. My daughter Christina is the joy of my life. She is smarter than I am, and, at age five, has a pretty solid understanding of
Acknowledgments—Continued

structural equation modeling. Jack seems less interested in statistics, but then, he is only two. Still, he communicates very clearly that he needs his daddy, but I understand that will not always be the case. Of course, this entire effort would be meaningless to me without them. I also know that we will be repaid the effort a thousand-fold in time.

Carl E. Hanssen
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................ ii

LIST OF TABLES ..................................................................................................... x

LIST OF FIGURES ................................................................................................... xi

CHAPTER

I.  INTRODUCTION ............................................................................................... 1
   Problem Statement ............................................................................................. 1
   Background .......................................................................................................... 3
      Multisite Evaluation ......................................................................................... 3
      The Advanced Technological Education Program ....................................... 5
      Theory-Based Evaluation ............................................................................... 7
      Combining Multisite and Theory-Based Evaluation ....................................... 7
      Structural Equation Modeling ........................................................................ 9
   Research Questions ............................................................................................ 10
   Relevance of the Study for Evaluators ............................................................ 10
   Definitions ......................................................................................................... 11

II.  REVIEW OF LITERATURE .............................................................................. 14
   Introduction ....................................................................................................... 14
   Multisite Evaluation ......................................................................................... 14
      Definition ........................................................................................................ 14
   Multisite Evaluation Challenges ..................................................................... 22
   Multisite Evaluation Examples ....................................................................... 23
<table>
<thead>
<tr>
<th>CHAPTER II. REVIEW OF LITERATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Advanced Technological Education Program</td>
</tr>
<tr>
<td>The ATE Program Evaluation</td>
</tr>
<tr>
<td>Theory-Based Evaluation</td>
</tr>
<tr>
<td>Definition</td>
</tr>
<tr>
<td>Theory-Based Evaluation and Multisite Evaluation</td>
</tr>
<tr>
<td>Structural Equation Modeling</td>
</tr>
<tr>
<td>Definition</td>
</tr>
<tr>
<td>SEM Examples in Evaluation</td>
</tr>
<tr>
<td>Relevance to This Study</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER III. METHODOLOGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
</tr>
<tr>
<td>Instrumentation</td>
</tr>
<tr>
<td>Measures</td>
</tr>
<tr>
<td>ATE Structural Equation Model</td>
</tr>
<tr>
<td>Sample</td>
</tr>
<tr>
<td>Data Preparation</td>
</tr>
<tr>
<td>Analysis</td>
</tr>
<tr>
<td>Specification</td>
</tr>
<tr>
<td>Identification</td>
</tr>
<tr>
<td>Estimation</td>
</tr>
</tbody>
</table>
# Table of Contents—Continued

## CHAPTER

### III. METHODOLOGY

- Evaluation and Modification ....................................................... 67
- Summary .................................................................................... 72

### IV. RESULTS .......................................................................... 74

- Full ATE Program Model ............................................................ 75
  - Overall Model Fit .................................................................. 75
  - Detailed Model Fit .................................................................. 76
  - Conclusions ......................................................................... 81

- Comparative ATE Program Models ............................................. 82
  - Collaboration-Impact (CI) Model ........................................ 82
  - Collaboration-Impact No Materials Development (CI no MD) Model .................. 83
  - Full Model No Materials Development (FM no MD) Model .......... 85
  - Comparative Fit .................................................................... 86

- Summary .................................................................................... 87
  - Fit of the Full ATE Logic Model ........................................... 87
  - Program Characteristics and Outcomes .................................. 88
  - Alternative Models ............................................................... 89

### V. DISCUSSION ......................................................................... 90

- Summary .................................................................................... 90
- Limitations ................................................................................. 93
Table of Contents—Continued

CHAPTER

V. DISCUSSION

Sample Size................................................................................ 93
Data Reliability............................................................................ 94
Measurement Models .................................................................. 95
Data Types ................................................................................ 98
Limitations of the ATE Program Theory..................................... 100
Summary of Limitations ...............................................................101
Implications.................................................................................. 102
Multisite Evaluation......................................................................102
Theory-Based Evaluation ...........................................................105
Structural Equation Modeling ......................................................108
Summary .....................................................................................111
Recommendations....................................................................... 112
Recommendations for Evaluators ...............................................112
Recommendations for Researchers .............................................114
REFERENCES................................................................................ 116

APPENDICES

A. HSIRB Approval Letter ............................................................ 123
B. Data Use Permission Letter .................................................... 125
C. ATE Evaluation Indicators ...................................................... 127
Table of Contents—Continued

APPENDICES

D. Survey Instrument .......................................................... 142
E. Measures Detail .......................................................... 186
F. Covariance Matrix ....................................................... 194
G. PRELIS and LISREL Code............................................. 196
LIST OF TABLES

1. Multisite evaluation examples ................................................................. 24
2. Normative and descriptive program models ............................................. 36
3. ATE program constructs ........................................................................... 54
4. Descriptive statistics for ATE program measures (n=115) ....................... 62
5. Standardized parameter estimates for measurement models 
   (full ATE program model) ....................................................................... 79
6. Comparative model fit ............................................................................ 86
7. Satorra-Bentler scaled chi-square difference tests .................................. 86

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## LIST OF FIGURES

1. ATE program model ................................................................. 8
2. ATE program structural model ................................................. 9
3. Initial ATE evaluation logic model ........................................... 31
4. Logic model for reducing alcohol related motor vehicle accidents and deaths ........................................... 35
5. IMD evaluation framework ..................................................... 41
6. ATE program structural model ................................................. 52
7. Full ATE program structural equation model ......................... 59
8. ATE program model with standardized path coefficients ........ 77
9. Stem-Leaf plot of standardized residuals .............................. 80
10. CI model with standardized path coefficients ....................... 82
11. CI no MD model with standardized path coefficients ........... 84
12. FM no MD model with standardized path coefficients .......... 85
CHAPTER I

INTRODUCTION

Problem Statement

Evaluation is the process of establishing the merit, worth, and value of things, and evaluations are the products of that process (Scriven, 1991). Two fundamental purposes exist for most program evaluations: (1) to document program results and/or (2) to improve programs (Scriven, 1991). Program evaluators recognize these purposes as summative and formative evaluation. The fundamental questions related to these purposes are

1. What happened as a result of the program?
2. How can the program be improved?

The answer to the first question is often obvious and can be answered by observing programs, collecting data, and documenting results. The answers to the second question are often less obvious, but far more compelling and, I would argue, more important. Ultimately, answering the question of how a program can be improved requires the evaluator to understand and report on why the program results that were observed occurred.

This study was concerned with formative evaluation and defining an approach for answering the question of why program results occurred. To

---

1 All references in this dissertation follow APA style as expressed in the American Journal of Evaluation.
accomplish this, this study integrated multisite evaluation with theory-based evaluation and structural equation modeling (SEM).

Multisite evaluation was chosen as the general context for this study because of the increasing number of large federally-funded evaluations of multisite programs that have been initiated over the past decade (Lawrenz & Huffman, 2003), and the author's general experience and familiarity with this type of evaluation. Within multisite evaluation, the specific context for this study was the National Science Foundation's (NSF) Advanced Technological Education (ATE) program. The evaluation of ATE produced a database that was used to describe program results. A descriptive program theory (Chen, 1990) of how ATE operated (Hanssen, et. al, 2003) also existed. SEM provided the analytical tool that was used to model the ATE program theory. The integration of these elements provided the basis for answering the question of why program outcomes occurred, which supports providing formative evaluation feedback. This is a step beyond solely answering the question of what occurred, which serves summative evaluation purposes.

The remainder of this chapter provides (a) the background for this study, (b) the research questions that guided this work, (c) the relevance of this study to the field of evaluation, and (d) definitions. Chapter II contains a review of literature that focused on the three central concepts in this study—multisite evaluation, program theory evaluation, and structural equation modeling. Chapter III outlines the methodology for this study; Chapter IV contains responses to the research questions. Chapter V concludes with a discussion on issues related to multisite evaluation, program theory evaluation, and SEM applications. Finally, recommendations for evaluators and researchers are provided.
Background

Multisite Evaluation

The distinguishing feature of multisite programs is that the program design specifically intends for the goals and objectives to be reached through funding multiple, independent projects\(^2\) (Greenberg, et. al., 2003; Hamilton, et. al., 2003; Worthen, et. at., 1997). Multisite programs have been categorized as controlled or uncontrolled (Sinacore & Turpin, 1991). Controlled implementation means the program has been implemented in the same way across all the projects. Uncontrolled implementation does not required uniform implementation at project sites. Multisite programs have often deferred to uncontrolled implementation because of the desire to learn what works under a variety of conditions or to compare results from competing program models (Greenberg, et. al, 2003; Sinacore & Turpin, 1991; Tushnet, 1995).

A typical federal multisite program might have the following characteristics: the program is established within a federal agency with broad objectives; the agency provides funding to numerous independent organizations that carry out program activities (i.e., projects); and each project that receives funding has a degree of discretion over its specific goals and objectives, and related activities. Thus, a program can be described as a collection of projects that seek to meet a defined set of goals and objectives (National Science Foundation, 2002). The result of this typical program design is that no two projects are the same, engage in the same

\(^2\)Throughout this dissertation, the term "program" refers to the overarching funding initiative; the term "project" refers to the individual organizations that have been funded by the program.
activities, nor produce, or are expected to produce, the same results. Evaluation of programs in this context depends on examining the program's separately funded projects.

Evaluators have faced significant challenges when evaluating multi-site programs (Sinacore & Turpin, 1991; Herrell & Straw, 2002). These challenges include addressing substantial variation in program implementation across sites, formulating and collecting consistent data on appropriate outcome measures, and synthesizing findings across program sites (Hamilton, et. al., 2003). Combined, these issues have made it difficult for evaluators to establish the merit and worth of an entire program based on information provided by individual projects.

Multisite evaluations may be prospective or retrospective (Herrell & Straw, 2002). Prospective multisite evaluation occurs when the multisite evaluator is involved in gathering data for the evaluation; retrospective multisite evaluations occur when the evaluator relies on data collected by the projects. The latter form is less common (Sinacore & Turpin, 1991) than prospective multisite evaluation.

In prospective multisite evaluations, a concern when collecting data has been gathering consistent information across projects. Two types of data have been commonly sought to support multi-site evaluations: (1) information about organizational characteristics and (2) evidence of results. It is common, however, that different projects establish different outcome measures, track outcomes in different ways, or that the measures available across projects are not appropriate for tracking program outcomes (Hamilton, et. al, 2003).
The Advanced Technological Education Program

The Advanced Technological Education (ATE) program grew out of a national interest in, and concern for, a balanced approach to developing and using technology to meet the nation’s work force needs (The Evaluation Center, 2000). In 1992, Congress passed the *Scientific and Advanced-Technology Act of 1992* (PL 102-476), which called for NSF to establish a national program to improve the education for technicians in advanced technology fields. The Act was intended to serve the ultimate goal of improving the competitiveness of the U.S. in international trade by increasing the productivity of the nation’s industries. This was to be accomplished by increasing the pool of skilled technicians in *strategic* advanced-technology fields. ATE was initiated by NSF in response to the Congressional mandate.

ATE’s goal was to expand the number and quality of skilled technicians in strategic advanced-technology fields. Specifically, ATE sought to accomplish this through several program areas, which were (a) developing model instructional programs, (b) providing professional development for faculty in advanced-technology fields, (c) establishing innovative partnership arrangements, (d) acquiring and implementing state of the art instrumentation, and (e) developing and disseminating instructional materials.

ATE made its first grants in the summer of 1994. As of 2003, nearly 550 awards had been made, totaling $300 million (National Science Foundation, 2003a). In Fall 2003, the program had approximately 230 active grants (National Science Foundation, 2003b).
The Evaluation Center at Western Michigan University was awarded a grant to conduct the external program evaluation of ATE. The evaluation is based on four questions (The Evaluation Center, 2003): (1) to what degree is the program achieving its goals? (2) is [the program] making an impact and reaching the individuals and groups intended? (3) how effective is [the program] when it reaches its constituents?, and (4) are there ways the program can be significantly improved?

Questions 1-3 addressed the question of what is occurring. ATE program goals were clearly defined and the evaluation examined the degree to which projects were productive within the primary program areas. The evaluation examined the degree to which projects are impacting students by gathering data about the number of program participants and the number of individuals placed in technician positions. The evaluation addressed effectiveness through targeted studies that explored each of the major program areas.

Question 4 partially addressed the question of why program results were occurring. The question of how the program can be improved implied the need to understand what contributed to program success. However, the specific question of understanding what did and did not work to produce desired results was not directly addressed by the evaluation.

The evaluation team engaged in a number of activities to answer these questions, including site visits with projects and business and industry partners, evaluation of instructional materials, regular meetings with NSF program staff, participation in conferences, and an annual survey. The survey addressed each of the ATE program areas, and projects that were active for at least one year were asked...
to report on their activities. Survey data were compiled in the ATE program evaluation database, which was used in this study.

Theory-Based Evaluation

The terms program theory evaluation (Bickman, 1990), theory-based evaluation (Weiss, 1997), and program logic (Funnell, 1997) have been used to describe the practice of basing a program evaluation on a causal model. The program theory/causal chain/logic model has provided a rationale for understanding how the program intended to produce its desired outcomes, thus enabling the evaluator to answer to the question of why program outcomes occurred. In the absence of validated program theory (which is common in education and human services), the foundation that has typically been used in theory-based evaluation is a model of the program's logic.

A program logic model can provide evaluators with a basis for answering the question of why outcomes occur. The core principal of theory-based evaluation has been to use the program's underlying assumptions or logic model to guide the program evaluation (Rogers, et. al, 2000). Within many program logic models, a series of intermediate antecedents and outcomes have been defined. The definition of antecedents and outcomes, or events and responses (Lipsey, 1993) provides a basis for measurement which can contribute to evidence about why program results occur.

Combining Multisite and Theory-Based Evaluation

Adopting the core principal of theory based evaluation, which is to use an underlying program theory as a guide for the program evaluation, when conducting multisite evaluation can help evaluators (a) describe the program and its results, and
(b) explain the reasons for those results, i.e., respond to both summative and formative evaluation purposes.

Figure 1 depicts the ATE program model, which is keyed to the program areas listed above. The figure depicts the program logic as follows: NSF provided support to projects with various characteristics; the organizational practices employed by these projects were the building blocks for program-related activities; and collaboration, materials development, and professional development efforts were expected to serve program improvement and directly impact the workforce through better-educated students (Hanssen, et. al., 2003).

![ATE program model diagram]

Figure 1. ATE program model
Lastly, the annual ATE evaluation survey was keyed to these program areas. As a result, data were available to examine the relationships between each program area. This provided an opportunity to attempt to answer the question of why program results occurred.

**Structural Equation Modeling**

The ATE program model outlined in Figure 1 implied a series of relationships that could be examined using SEM. SEM tests the relationships between defined constructs and the associated theory and specifies how constructs are related. To answer the question of "why" program results occurred within the ATE program, it was necessary to construct a structural model based on the constructs that defined the program. Figure 2 presents an initial model based on the seven ATE constructs depicted in Figure 1.

![ATE program structural model](image)

**Figure 2. ATE program structural model**

The ATE program structural model followed directly from the program logic model, which was based on the program design. The structural model was described as follows: collaboration was dependent on project characteristics and organizational practices; results in the three work categories (1) materials development, (2) professional development, and (3) program improvement were
dependent on collaboration; professional development and program improvement were also dependent on materials development; program improvement was also dependent on professional development; student impact was dependent on program improvement.

Research Questions

The following questions were addressed in this study using the ATE program evaluation database. Answers to these questions provided guidance to program evaluators faced with the challenge of providing formative evaluation evidence, i.e., answering the question of why outcomes occur.

In the context of multisite evaluation,

1. To what extent does the ATE program logic model, in components or in its totality, fit the empirical data?

2a. If the model, in components or in its totality, can be fitted, what are the dynamics with which the programmatic characteristics contribute to program outcomes?

2b. If the empirical fit of the model is not optimal, can the current logic model be modified to improve the fit?

Relevance of the Study for Evaluators

The emergence of multisite evaluation as a focus for program evaluators was highlighted in two *New Directions for Evaluation* volumes edited by Sinacore and Turpin, and Herrell and Straw in 1991 and 2002, respectively. These volumes provided definitions and examples of multisite evaluations and in many ways...
launched this stream of evaluation thought. More recently, Lawrenz and Huffman (2003) discussed the participatory nature of multisite evaluations in an article published in the *American Journal of Evaluation*. What is notable about this recent article, however, is that the five programs used as examples in the discussion—all of which were NSF programs and one of which was the ATE program—represented literally hundreds of millions of dollars in federal program funding.

The size and scope of these programs underscores the importance of developing innovative methods and applying sound practice in evaluating these efforts. Similarly, evaluations of these programs should be comprehensive in that they should address both summative and formative issues.

Theory-based evaluation and structural equation modeling are both widely accepted social science research tools. Combined, these approaches represent an opportunity to innovate an approach to addressing formative evaluation within a multisite program context. While these tools and this approach may not be appropriate in all multisite program contexts, they may help evaluators provide additional evidence to support evaluation findings and recommendations. Given this, evaluators who evaluate multisite programs should be interested in this approach.

**Definitions**

**ATE.** The Advanced Technology Education program funded and administered by the National Science Foundation.

**Controlled implementation.** Implementation of a program where the program administrators define the specifics of how the program is implemented, resulting in decreased variation in implementation between projects.
Evaluation. The process of determining the merit, worth, or value of some object or the product (i.e., the report) of that process.

Formative evaluation. Evaluation that is focused on improving programs.

Multisite evaluation. Evaluation of a multisite program that looks across funded projects, aggregates results, and establishes merit and worth of the program based on the aggregated data.

Multisite program. A program that is implemented through the funding of multiple projects at different locations.

NSF. The National Science Foundation, which funds and administers the ATE program.

Program. A funding initiative that targets a specific, expressed need.

Program theory. The theory or logic that supports the need for and design of a program; program theory suggests a series of causal relationships that contribute to the ultimate program outcome (i.e., if this occurs then this happens, then this results).

Project. An entity or organization that receives funding under a program.

Prospective evaluation. Evaluation of a multisite program where the central evaluator is directly involved in gathering data used in the evaluation.

Retrospective evaluation. Evaluation of a multisite program where the data used in the evaluation is developed by the individual projects and then given to the central evaluator.

Structural Equation Modeling. The process of defining relationships between latent variables (constructs) based on observable data.
**Summative evaluation.** Evaluation that is focused on documenting program results.

**Theory-based evaluation.** An evaluation that is guided by the underlying program theory or logic that defines the program.

**Uncontrolled implementation.** Implementation of a program by funded projects that is not controlled or pre-defined by the program administrator.
CHAPTER II

REVIEW OF LITERATURE

Introduction

The review of literature explored the three central concepts in this dissertation: (1) multisite evaluation, (2) theory based evaluation, and (3) structural equation modeling. The discussion of each concept provides a definition (or competing definitions) and examples of how the concept has been described and used in past work. As a conclusion, the relevance and significance of past work to this dissertation is discussed.

Multisite Evaluation

Definition

Many evaluators and researchers acknowledge that multisite evaluations are common phenomena (Sinacore & Turpin, 1991; Tushnet, 1995; Worthen, et. al., 1997; Herrell & Straw, 2002). Various terms have been used to describe multisite evaluation, including cluster evaluation (W.K. Kellogg, 2000), cross-site evaluation, multisite evaluation, and multi-center clinical trials (Herrell & Straw, 2002). While these variants share common characteristics, there is considerable debate about the characteristics that differentiate these types of multisite evaluations.

An important component of multisite evaluation, in any of the forms listed above that is not explicitly stated in the literature (though strongly implied), is that a
multisite evaluation is coordinated/conducted by a single evaluator or evaluation team. This is an important distinction because many programs that are implemented at multiple locations require each location to also engage their own evaluator. These evaluators are not conducting a multisite evaluation, though they may participate in, or provide data for, the multisite evaluation. Throughout the remainder of this discussion, the term evaluator will be used to describe the multisite evaluator, rather than a specific site evaluator.

An initial definition of multisite evaluation comes from Sinacore and Turpin (1991). In fact, some authors credit them with coining the term “multisite evaluation,” or MSE (Herrell & Straw, 2002; Tushnet, 1995). Sinacore and Turpin argue that the distinguishing feature of an MSE is “its implementation at different sites with an analysis of original data” (1991, p. 7). The term "original data" is significant in that it implies that the evaluator uses data that has been gathered expressly for the purposes of the multisite evaluation. Sinacore and Turpin (1991) further differentiate multisite evaluation as either prospective or retrospective. Prospective evaluation occurs when the evaluator intends to use multiple sites at the beginning of the evaluation. Retrospective evaluation occurs when the data from different evaluations on a similar topic are combined and form the basis for the multisite evaluation. Sinacore and Turpin’s definition of retrospective appears to contradict the earlier statement that a multisite evaluation relies on “original data,” thus supporting the argument that an evaluation can not be a multisite evaluation unless it is prospective.

Herrell and Straw (2002), however, provide alternative definitions of prospective and retrospective evaluation that are more useful. They assert that a prospective evaluation is one where the multisite evaluator is actively involved in the
collection of data; conversely, a retrospective evaluation relies on data already collected by the sites. Unfortunately, Herrell and Straw are mute on the issue of the multisite evaluator defining measures and then relying on the sites to collect the data. For the purposes of this discussion, however, we will consider this to be form of a prospective evaluation. Thus, a multisite evaluation can clearly be prospective or retrospective, irrespective of the intent (as described by Sinacore and Turpin) of the evaluator at the outset. The definitions provided by Herrell and Straw seem more relevant and are used throughout the remainder of this discussion.

Sinacore and Turpin (1991) also classified MSEs according to the implementation of the program being evaluated. They indicated that programs are implemented either in a controlled or uncontrolled manner. Examples of controlled multisite program implementation are often cited in the healthcare literature, one form of which is multicenter clinical trials (Herrell & Straw, 2002). Uncontrolled implementation, however, appears much more frequently in the literature when talking specifically about multisite evaluation such that Tushnet (1995) argued that uncontrolled program implementation is a distinguishing feature of multisite evaluations.

Building on this discussion of prospective and retrospective evaluation, and controlled and uncontrolled program implementation, Herrell and Straw (2002), distinguish between three general variants of multisite evaluations: (1) cluster evaluation, (2) multisite evaluation, and (3) multicenter clinical trials. According to Herrell and Straw, cluster evaluations tend to be exploratory in nature the primary purpose of examining variation in program implementation. Thus, these evaluations may be prospective or retrospective but the program being examined is usually
implemented in an uncontrolled manner. Conversely, multicenter clinical trials tend
to be confirmatory with the purpose of estimating the impact of treatments. Program
implementation in these instances is tightly controlled and the evaluation is
necessarily prospective to ensure that the data collected can support the
confirmatory purpose of the evaluation.

Using Herrell and Straw's classification scheme, multisite evaluations fall in
the middle of the continuum between cluster evaluation and multicenter clinical trials
and may adopt the distinct characteristics of either or both of these variants. Thus, a
MSE may be either exploratory or confirmatory with the purpose of estimating impact
and examining variation. Similarly, MSE may be prospective or retrospective and the
program implementation being evaluated may be controlled or uncontrolled.

To summarize, multisite evaluation can be defined as an evaluation that is
conducted by a central evaluator and that uses data collected from each location.
The data used for the evaluation may be collected by the evaluator (prospective) or
compiled by the sites and then provided to the evaluator (retrospective). The
program being evaluated may be implemented in a controlled or uncontrolled
manner. It is the issue of the program that is addressed next, followed by a more
detailed discussion of the evaluation purposes.

Objects of Multisite Evaluation

A substantive discussion of what constitutes the evaluand, or object of
evaluation, in a multisite evaluation is missing from the literature. There is an implicit
assumption that the evaluand must be a defined program. Thus, when authors
describe multisite evaluations, they are speaking about the evaluation of a program
that is implemented at multiple sites, ala Sinacore and Turpin. Lawrenz and Huffman
(2003) demonstrate, by describing five National Science Foundation programs, that a key feature of multisite programs is that there is some tangible element, i.e., a program or issue, that links all of the sites together.

Greenberg, et. al. (2003), present an instructive example of how evaluating responses to an overarching problem can be conceptualized as a multisite evaluation that examines results from different welfare to work programs. This example satisfies the basic definition of multisite evaluation—data from multiple sites are used in the evaluation by a central evaluator. This example can also clearly be categorized as uncontrolled program implementation and retrospective evaluation. However, the object of the evaluation is not a multisite program, rather it is an overarching response to a defined problem. This response led to the creation of multiple programs, and evidence about how the problem was addressed was found using data that were collected at each of the locations were the response was exhibited.

This example supports the idea that what links sites together in a multisite evaluation is not necessarily a "program," but rather an problem being addressed. What is important is that each site is implementing an intervention to address that issue. The interventions may exist within the context of a multisite program, or they may exist in response to a defined issue. An issue could be articulated as a policy statement or by a belief in the need to address some social problem. Within the context of the ATE program, the issue being addressed was a dearth of qualified technicians to support U.S. industry needs (The Evaluation Center, 2000). The example cited by Greenberg, et. al., (2003) identifies the issue as a recognized problem of transitioning welfare recipients to work. The responses manifested at the
sites in Greenberg's study all targeted this issue, though they did not operate within a coordinated programmatic context.

The welfare to work example helps clarify (and confuse) the issue of multisite evaluation. In addition to the summary definition provided above, multisite evaluations may target issues and/or problems, not just defined programs. Thus, an important factor in conducting multisite evaluation in these cases is the availability of consistent data across sites, in cases where the evaluation is not linked to a single program effort.

Multisite Evaluation Data

In the Greenberg, et. al. (2003) example, the consistent data that were available fell into two categories—outcome measures and site characteristics. The primary outcome measure for the welfare to work programs was the increase in income for program participants. Site characteristics included demographics, unemployment rates, and various program related activity measures (Greenberg, et. al., 2003). From the outcome measures, an effect size for each site was determined. This effect size was then used as the unit of analysis in aggregating program impacts and comparing program effects size across sites.

This example defines two additional characteristics of multisite evaluation, which help clarify the concept. First, the data collected by the central evaluator from various sites (whether prospective or retrospective) must be consistent. Secondly, the data from multiple sites should be appropriate for pooling to allow for the determination of an overall effect (i.e., program impact) as well as comparisons across sites. Even cluster evaluation proponents, who view the primary function of multisite evaluation as exploratory (Herrell & Straw, 2002), seem to adhere to this
concept in their activities by defining and collecting at least a minimal set of data across sites for determining an overall effect and to allow comparisons of results across sites (W.K. Kellogg, 2000). Furthermore, Kalaian (2003) argues that in cases where consistent data are not available across sites, multisite evaluation is actually more akin to meta-analysis (Glass, McGaw, & Smith, 1981) and should not be considered multisite evaluation.

**Multisite Evaluation Purposes**

The definitional elements of multisite evaluation imply that multisite evaluation has a broad array of purposes. For example, Herrell and Straw (2002) distinguish between an exploratory purpose and a confirmatory purpose. The critical exploratory question is determining what works under a variety of conditions. The primary confirmatory question is understanding what happened, i.e., did the intervention work. These purposes are implicitly aligned with formative and summative evaluation, though an evaluation may be exploratory, but would only be truly formative if the findings were used to improve program implementation. Overall, it is clear that multisite evaluation may address both the question of what occurred in a program as well as why outcomes were produced.

Worthen, et. al, (1997) provide a list of potential multisite evaluation purposes which are consistent with the above discussion. They are

1. To determine the overall effect of the program, when effects are aggregated across all program sites,

2. To determine whether the program works under the variety of conditions and circumstances that exist where it has been implemented,
3. To determine how the program interacts with specific site characteristics (e.g., demographic differences in program participants, varying placements of the program within agencies governing structures),

4. To monitor individual project compliance in implementing the program according to its specifications or standards,

5. To compare program performance across projects to identify the most effective and ineffective ways of operating the program,

6. To determine which projects should be continued or discontinued in the program, and

7. To share effective practices, lessons learned, and other insights gained in one project that could be beneficial to other projects.

Finally, Tushnet argues that the primary purpose for conducting a multisite evaluation is to understand what works within a program under different conditions and implementations. Multiple authors (Greenberg, et. al., 2003; Worthen, et. al., 1997; Sinacore & Turpin, 1991, W.K. Kellogg, 2000) agree with this argument and cite this comparative element as the primary function of multisite evaluations, and, indeed, as one of its primary benefits. This comparative purpose speaks directly to the issue of why outcomes occur.

Summary

The definition of multisite evaluation presented above is summarized as follows:

1. The object of a multisite evaluation is defined primarily by a problem and a tangible response to that problem.
2. The response to the problem is evident at multiple locations, which is the critical linkage across the sites.

3. Multisite evaluations are conducted by a central evaluator who examines data from multiple locations.

4. Two types of data are required (1) information on results and (2) site characteristics.

5. Data collected are consistent in that they support pooling for determining (1) overall effects and (2) comparisons across program sites.

Multisite Evaluation Challenges

Evaluators have faced significant challenges when evaluating multisite programs (Sinacore & Turpin, 1991; Herrell & Straw, 2002). These challenges have included (a) addressing substantial variation in program implementation across sites, (b) formulating and collecting consistent data on appropriate outcome measures, and (c) synthesizing findings across program sites (Hamilton, et. al., 2003).

In addition, there are a myriad of other issues to consider when examining multisite evaluations. These issues fall into two categories—program related issues and methodological issues.

Program issues include context (Tushnet, 1995), the definition of a program, the role of the administrator (W.K. Kellogg, 2000), the variation in program goals and objectives (Hamilton, et. al., 2003), the definition of a “site” (Herrell & Straw, 2002), and the flexibility of different sites to engage in different activities (Lawrenz & Huffman, 2003; National Science Foundation, 2003). Thus, multisite evaluators must precisely address each of these issues to ensure that the object of the evaluation is clearly understood by the evaluation stakeholders.
Methodological issues include the use of quantitative versus qualitative data (Lamberti & Katzenmeyer, 1996), the purpose of the evaluation, i.e., formative or summative, confirmatory or exploratory (Herrell & Straw, 2002), sample size (Sinacore & Turpin, 1991), and the unit of analysis, i.e., site versus individual program participant (Greenberg, et. al., 2003). It is important to note that these methodological issues are not unique to multisite evaluation. For example, all evaluations must consider the use of quantitative versus qualitative data. In a multisite context, however, the use of qualitative data may detract from the ability to aggregate program effects across sites.

Several examples of multisite evaluations are provided below that highlight the variations in how different MSEs satisfy the five-point definition provided above. Additionally, other critical program and/or methodological issues are highlighted to demonstrate how variations in these issues do not disqualify an evaluation as a multisite evaluation. Lastly, this discussion concludes with a detailed examination of the ATE program as the specific context of this study.

**Multisite Evaluation Examples**

There are numerous examples of MSEs in the literature (Anderson, et. al., 2003; Greenberg, et. al., 2003; Hamilton, et. al., 2003; Barbor, et. al., 2002; W.K. Kellogg, 2000; Schleger, et. al., 1999; NSF, 1998; Tushnet, 1995). Table 1 presents five examples of multisite evaluation, one of which is the ATE program, which is the specific context for this study. These examples are highlighted because they represent variants of multisite evaluation practice, yet each satisfies the basic characteristics of a multisite evaluation. In addition, while these examples should not be taken to be systematically representative of all multisite evaluations, their
Table 1

Multisite evaluation examples

<table>
<thead>
<tr>
<th>Problem/Program(s)</th>
<th>Sites</th>
<th>Site Data</th>
<th>Results Data</th>
<th>Prospective/Retrospective</th>
<th>Controlled/Uncontrolled</th>
<th>Program Issues</th>
<th>Methodology Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place children in permanent homes (FFK) (W.K. Kellogg, 2000)</td>
<td>11</td>
<td>States, contextual data</td>
<td>Placement figures</td>
<td>Prospective—evaluators worked with site evaluators to define data</td>
<td>Uncontrolled—implementation was unique and diverse</td>
<td>Program designed to improve existing systems rather than establish a new system</td>
<td>Consistent data across sites difficult to obtain; purpose was primarily exploratory</td>
</tr>
<tr>
<td>Substance abuse prevention and early intervention (WMC) (Schlienger, et. al., 1999)</td>
<td>9</td>
<td>Program strategy; workplace characteristics</td>
<td>employee survey, utilization data, implementation descriptions</td>
<td>Prospective—evaluators collect survey data</td>
<td>Uncontrolled—implementation varied according to program strategy</td>
<td>All program goals not necessarily addressed by all sites</td>
<td>Multi-protocol, multi population—generate a wide body of less definitive knowledge</td>
</tr>
<tr>
<td>Increase # and quality of technicians in U.S. workforce (ATE) (Hansen, et. al., 2003)</td>
<td>100+</td>
<td>Institution, funding level, tech emphasis</td>
<td>Students impacted &amp; placed, courses developed</td>
<td>Prospective—evaluators collect data annually via survey</td>
<td>Uncontrolled—projects select from a variety of program activities</td>
<td>Variation in selected program activities</td>
<td>Summative evaluation w/ some formative elements</td>
</tr>
<tr>
<td>Move welfare recipients to steady jobs (Greenberg, et. al., 2003)</td>
<td>13</td>
<td>Program name; population demographics</td>
<td>Earnings of program and control groups</td>
<td>Retrospective—evaluators relied on existing data</td>
<td>Uncontrolled—2 different strategies; variation within strategy</td>
<td>Two different programs under consideration</td>
<td>Large samples within sites, small sample across</td>
</tr>
<tr>
<td>Marijuana dependence (MTP) (Babor, et. al., 2002)</td>
<td>3</td>
<td>Demographics of participants</td>
<td>Frequency of use during assessment period</td>
<td>Prospective—study variables defined up front prior to grants</td>
<td>Controlled—3 protocols used at each site</td>
<td>Training of therapists; program and evaluation are co-mingled</td>
<td>Human subjects considerations, i.e, denying treatment to control group</td>
</tr>
</tbody>
</table>

Note. The definitions of prospective and retrospective evaluation, and controlled and uncontrolled implementation are based on Herrell & Straw (2002).
characteristics are reasonably representative of the scope and breadth of many MSEs.

These examples show that by and large, most multisite evaluations consider a relatively small number of sites. The exception to this is the ATE multisite evaluation which gathers data from over 100 project sites annually (Hanssen, et. al., 2003). In the case of ATE, the unit of analysis for the evaluation is the funded project, whereas in other evaluations, the unit of analysis may be a program participant, resulting in larger samples. For example, the sample size for the MTP program was 450 (Barbor, et. al., 2002) because the unit of analysis was individual program participants. The sample size was in excess of 40,000 for the retrospective evaluation of welfare to work programs documented by Greenberg (2003) and his colleagues. Thus, the sample size and unit of analysis are important considerations for multisite evaluators (Sinacore & Turpin, 1991).

However, the critical factor here is that sampling unit used in the evaluation describes the unit of analysis, and the maximum attainable sample size for the evaluation. As a result, collecting data from multiple subjects at a given site does not necessarily improve sample size if the sampling unit for the evaluation is an individual site.

These examples also highlight the idea that most multisite evaluations are prospective and address uncontrolled implementation. (I make this statement understanding that multicenter clinical trials, which are a variant of multisite programs that use controlled implementation and prospective evaluation (Herrell & Straw, 2002) are widespread, but wanting to focus this discussion on the MSEs that fit neither the extremes of cluster evaluation or multicenter clinical trials.) This is
consistent with Tushnet's assertion (1995) that uncontrolled implementation was a key characteristic of multisite evaluation. Similarly, Sinacore and Turpin (1991) indicated that prospective evaluation was most common, even though their definition of "prospective" was later clarified by Herrell and Straw (2002) and is used differently than Sinacore and Turpin described.

Finally, both the FFK (W.K. Kellogg, 2000) and WMC (Schlenger, et. al., 1999) programs could be considered cluster evaluations. Both were examples of prospective evaluation and uncontrolled implementation. Each had, as a primary objective, the desire to create a large amount of new information. This new information could be used to answer the question of why outcomes occurred and might also be used in a formative manner to improve program design and implementation. In this sense, these evaluations were primarily exploratory, rather than confirmatory, which is a primary characteristic of the cluster evaluation variant (Herrell & Straw, 2002). Of course, this does not exclude other multisite evaluations from exploring variations in program implementation strategies, nor does this distinction warrant substantively differentiating cluster evaluations from other multisite evaluations.

The Advanced Technological Education Program

ATE grew out of a national interest in and concern for a balanced approach to developing and using technology to meet the nation's work force needs (The Evaluation Center, May 2000). In 1992, Congress passed the Scientific and Advanced-Technology Act of 1992 (PL 102-476), which called for NSF to establish a national program to improve the education for technicians in advanced technology fields. The Act was intended to serve the ultimate goal of improving the
competitiveness of the U.S. in international trade by increasing the productivity of the nation’s industries, which in turn was to be accomplished by increasing the pool of skilled technicians in strategic advanced-technology fields.

With respect to the definition of multisite evaluation outlined above, the Congressional action represented a specific response to the stated problem of a lack of qualified technicians in the U.S. As described below, ATE became one response to that problem, but it is possible that other responses were also initiated. Thus, the object of a hypothetical multisite evaluation could be the overall response to the stated problem of too few trained technicians as manifested in different educational initiatives (of which ATE was one).

Congress emphasized the role of two-year colleges for this program. As House Report 102-508, p. 4 stated, “Two-year colleges are a major contributor to higher education and have become the largest pipeline to postsecondary education in the United States. In 1990, 1350 two-year colleges enrolled approximately 5 million students, representing 43 percent of all undergraduate students and constituting 40 percent of all institutions of higher education. Approximately 30 percent of students enrolled in two-year colleges transfer to four-year colleges and universities.”

**ATE Program Objectives**

NSF initiated ATE in response to the Congressional mandate. ATE was created in the Education and Human Resources Directorate (EHR) and was co-managed by two Divisions, the Division of Undergraduate Education (DUE) and the Elementary, Secondary, and Informal Education Division (ESIE). Consistent with the Congressional mandate, ATE’s overarching goal was to expand the pool of
skilled technicians in strategic advanced-technology fields. ATE directed funding to community colleges, which was also consistent with the Congressional mandate, in pursuit of this goal.

The methods for achieving this goal, however, were varied, which led to, in terms of the multisite evaluation definition, uncontrolled program implementation. The type of activities in which funded project engaged were described by the specific ATE program goals listed below

1. Develop model instructional programs in advanced-technology fields.

2. Provide professional development of faculty and instructors in advanced-technology fields.

3. Establish innovative partnership arrangements that (a) strengthen the relationships between associate-degree-granting colleges and secondary schools in the communities, (b) build strong working relationship between the associate-degree-granting colleges and the businesses, industries, and other appropriate public and private sector entities that need skilled technicians in their work force, and (c) provide for private sector donations, faculty opportunities, etc.

4. Acquire and implement state of the art instrumentation.

5. Develop and disseminate instructional materials.

Over the course of the program, these goals became commonly described as (1) program improvement, (2) professional development, (3) collaboration, and (4) materials development (The Evaluation Center, 2000).
ATE grants were awarded in three categories: centers, projects, and articulation partnerships (NSF, 2003c). Centers received more funds and had a broader range of activities than did projects. Projects tended to focus on only one or two of the above objectives, though larger projects often engaged in all the program activities. Articulation partnerships focused on collaborative relationships between educational institutions at multiple levels to promote access to higher levels of education. Overall, ATE funded a relatively small number of centers (10-20), a large number of projects, and only a few articulation partnerships. However, the small number of articulation partnerships funded was primarily due to the fact that the first awards in this category were made in 1994 (Hanssen, et. al, 2003).

ATE Operations

Operationally, the ATE program was straightforward. The program solicited and reviewed preliminary proposals from institutions and their partners in or around April of each year. The preliminary proposals were reviewed and feedback provided to applicants, together with an overall judgment as to whether or not a full proposal was encouraged. The applicants, regardless of the ATE program feedback, were allowed to submit a full proposal, which was due in October of each year. Full proposals were reviewed and funded on a merit basis. Funded institutions then used the grant funds to conduct the proposed activities.

Program officers handled oversight of grant awards and the nature and extent of oversight varied. NSF hosted annual Principal Investigator meetings for all projects, which provided opportunities for interactions among projects and with program officers. Monitoring of small projects (i.e., projects receiving relatively small annual awards) depended largely on interactions between the program officer and...
the project director. Large projects (i.e., over $500,000 per year) and all centers were expected to create advisory boards or National Visiting Committees that reviewed grant activity and provided input to the principal investigator for the grant and to NSF.

ATE made its first grant awards in the summer of 1994. As of 2003, nearly 550 awards had been made, totaling $300 million (National Science Foundation, 2003a). In Fall 2003, the program had approximately 230 active grants (National Science Foundation, 2003b).

The ATE Program Evaluation

The Evaluation Center at Western Michigan University was awarded a grant in 1999 to conduct an external evaluation of the ATE program. The initial evaluation period was for three years and was extended, in 2002, for three additional years. NSF funded the evaluation for approximately $500,000 annually. The evaluation is based on four questions (The Evaluation Center, July 2003)

1. To what degree is the program achieving its goals?
2. Is [the program] making an impact and reaching the individuals and groups intended?
3. How effective is [the program] when it reaches its constituents?
4. Are there ways the program can be significantly improved?

Questions 1-3 focused on addressing the question of what is occurring, which is consistent with the summative focus of the evaluation. Question 4 partially addressed the question of why program results were occurring. The question of how the program can be improved implies the needs for an understanding of what contributed to program success. However, the specific question of understanding
what worked and what did not work to produce desired results was not an explicit purpose of the evaluation. However, consistent with the definition of multisite evaluation, the ATE program evaluation was interested in both documenting program results and in supporting program improvement.

**Program Evaluation Indicators**

During the first year of the evaluation, the evaluation team articulated a detailed list of indicators to guide data collection activities during the evaluation (The Evaluation Center, 1999a). These indicators were developed at three levels (see Appendix C): (1) the initiation and support of projects and centers, (2) development, i.e., short term outcomes and impacts, and (3) viability of center and project work. Each family of indicators defined drivers, enabling outcomes, short-term outcomes, and, in the case of the level 3 indicators, long term outcomes. In addition, the relationship between these indicators was articulated in a high-level program logic model as shown in Figure 3.

![Logic Model](https://example.com/ate_logic_model.png)

**Figure 3.** Initial ATE evaluation logic model

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Interestingly, the program logic model shown above was not explicitly used to guide the evaluation, especially in more recent years, i.e., 2003 to present. The value of this logic model appears to have been to articulate detailed questions to be used in various evaluation activities. It is apparent though, that the model, and its supporting indicator levels, was constructed with causal relationships in mind. The use of the terms "drivers," "short term outcomes," and "long term outcomes" implies a set of causal relationships that provide explanatory evidence about program results.

**Evaluation Activities**

The evaluation team engaged in a number of activities to answer the four broad evaluation questions. These activities included site visits to 13 high performing projects, site visits to business and industry partners, evaluation of materials produced by projects, regular meetings with NSF program staff, participation in PI conferences, and an annual web-based survey.

The annual survey was the central evaluation activity and provided a large amount of data for the evaluation. The survey employed a descriptive design in that its purpose was not to attribute causality but to describe the program (O'Sullivan & Rassel, 1995). The specific survey items (see Appendix D) addressed some of the level 1 and 3 indicators, but focused primarily on the level 2 indicators. These indicators examined results in each of program activity areas.

Each year, all ATE projects that were active for at least one year at the time the survey was administered (usually February through April) were selected to participate. Survey data were collected, analyzed, and reported each year (Gullickson, et. al., 2000, 2001, 2002; Hanssen, et. al., 2003). The annual survey report was primarily summative, but also provided overall recommendations for
improving ATE. In 2003, the report also contained a four-year summary of ATE program trends and introduced an updated program logic model that was based on the relationships between ATE program activities (see Figure 1 from Chapter I). This program logic model was significant in that it re-opened the opportunity to establish causal relationships between various program components.

**Evaluation Recommendations**

In response to the fourth evaluation question, the 2003 annual survey report (Hanssen, et. al, 2003) made two recommendations for program improvement. The first concerned the use of evaluation by projects and argued for NSF to encourage more substantive use of evaluation within projects to guide project activities. This recommendation was based on findings that few projects utilized evaluation throughout the development and implementation of materials development and professional development activities (Hanssen, et. al., 2003). The second recommendation concerned the recruitment and retention of minority and traditionally underserved populations. The report encouraged NSF to seek new methods for recruiting minority students into ATE-sponsored programs to broadly disseminate these ideas to all ATE projects. This recommendation was based on the findings that minority representation had not improved over time across ATE-funded projects (Hanssen, et. al., 2003).

The recommendations provided in the annual survey report, were examples of formative evaluation in that there were specifically targeted at improving the ATE program. Despite this, the report did not provide any evidence to show why adopting these recommendations would improve program results in terms of improving the number and quality of technicians in the workforce. In other words, there was no
justification or explanation for the formative feedback, i.e., the evaluation did not answer the question of why outcomes were produced.

The evaluation of ATE has done an excellent job of describing what happened in the ATE program, but has not adequately describe why those results occurred, which is a key concern of multi-site evaluations (Worthen, et. al., 1997). The program evaluators have not focused on what activities and/or conditions contributed to results produced by the projects. For example, was the level of funding received by a project related to the number of students who were impacted by new instructional programs? Or, what was the contribution of funding to student participation levels versus other organizational factors, such as participation in monitoring activities or the age of a project? Answers to these types of questions get to the heart of the issue of why results occur and provide the fundamental basis for this study. The application of a program logic model can help evaluators answer these types of questions in a multisite context.

Theory-Based Evaluation

Definition

The terms program theory evaluation (Bickman, 1990), theory-based evaluation (Weiss, 1997), and program logic (Funnell, 1997) have been used to describe the practice of basing a program evaluation on a causal model. For the purposes of this study, the term theory-based evaluation will be used to encompass all of these variants, as each of these variants suggests a program evaluation process that focuses on (a) the most effective program components, (b) the mediating causal processes through which a program works, and (c) the
characteristics of the participants, service providers, and context that moderate the relationships between a program and its outcomes (Donaldson, 2003).

A causal model is an important tool for theory-based evaluation because such a model provides the rationale for how a program produces its results. In the absence of a tested program theory (which is common in education and human services), the foundation that has typically been used in theory based evaluation is a program logic model. The ATE program models presented in Figures 1 and 3 above are two examples. Figure 4 below presents a third example.

![Logic model for reducing alcohol related motor vehicle accidents and deaths](image)

*Figure 4. Logic model for reducing alcohol related motor vehicle accidents and deaths*

The model in Figure 4 depicts a single overarching program goal—to reduce premature death and preventable injuries—but it also contains several intermediate measurable objectives that, while ends in themselves, contribute to the overall program goal. For example, eliminating drinking and driving in alcohol-dependent individuals should, as indicated, reduce alcohol related MVA fatalities, which should,
in turn, reduce premature death and preventable injuries. Thus, each intermediate outcome has antecedents and is an antecedent for downstream program results.

**Normative and Descriptive Program Theory**

Chen (1990a; 1990b) described two primary types of causal models—normative and causative (p. 43). A normative model describes how a program is intended to work. A causative model shows how the program actually works. Rogers (2000) applied the terms *normative* and *descriptive* models to Chen's normative and causative models. These terms seem more appropriate given the ensuing discussion and thus are used throughout this study.

As shown in Table 2, there are three issues that help distinguish between normative and descriptive models. They are (1) when the model was developed, (2) how it was developed, and (3) how it was used in an evaluation. Each of these issues has implications for theory based evaluation.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normative and descriptive program models</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Issue</th>
<th>Normative</th>
<th>Descriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>When</td>
<td>During program development; prior to the evaluation; at start of the evaluation</td>
<td>After the program has been operating; during the course of the evaluation</td>
</tr>
<tr>
<td>How</td>
<td>Deductively based on hypothesized relationships</td>
<td>Inductively based on observed program operations</td>
</tr>
<tr>
<td>Use</td>
<td>Confirmatory (summative)</td>
<td>Confirmatory (summative) and exploratory (formative)</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The issue of when the program model is developed is notable because there is potential overlap between the timing of the model development and the type of model developed. The important factor is that a normative theory describes how a program is intended to operate and is most likely to be articulated during program development or, at a minimum, at the very beginning of an evaluation. Conversely, a descriptive program model can not be developed at the start of the program because it is based on how the program actually operates. Thus, a descriptive model is most likely to be developed after a program has been in operation.

Rogers, et. al., (2000), point out that program theory may be developed ahead of the evaluation or during the course of the evaluation. In the former case, the implication is that the evaluation would be focused, in a summative manner, on testing the program theory, i.e., evaluating the normative model. Thus, the evaluator may be in a position to determine if the theory breaks down or if the implementation of the program breaks down (Lipsey, 1993). The analogy to the multisite evaluation is that a normative program model would support prospective program evaluation that has primarily a confirmatory purpose (e.g., a multicenter clinical trial).

Program theory developed during the evaluation would result in a descriptive model, and suggests an inductive approach to developing the program theory. In this case, the evaluators, in the course of examining program processes and results, articulate a model of how they believe the program works. The analogy to multisite evaluation is an evaluation with an exploratory purpose (e.g., a cluster evaluation). In future evaluations, this model may function as the apriori program theory which would be tested in the manner Lipsey describes, but in the present evaluation, it can only serve as the observed case of how the program is operating at that time.
Overall, the distinction between Chen's normative and descriptive causal models primarily concerns when the causal model was developed, how it was developed, and how it is used in a theory based evaluation. In the theory-testing case using a normative model, measures, data, and analyses are concentrated on the task of testing the underlying relationships expressed by the program theory (Bickman et. al., 1998). In the inductive case, measures, data, and analyses are focused on describing what has occurred in the program by using a descriptive program theory.

Applying the multisite evaluation terminology described earlier, use of a normative model might be termed prospective, while using a descriptive model could be termed retrospective. Despite this obvious link, I am hesitant to classify all uses of a descriptive model as retrospective evaluations. Recall that Herrell and Straw (2002) defined retrospective evaluation as relying on existing data where the evaluator did not have a role in defining measures or collecting data. The reverse is true in the prospective case, but that does not exclude prospective evaluation from using a descriptive program model. In fact, it is likely that the development of the descriptive model was directly the work of the evaluator, working from data that he/she collected. Similarly, a prospective evaluation may begin after a program has been in operation for several years, as long as the evaluator has a clear role in defining measures and collecting evaluation data. This latter case does not require the existence of a normative program model.

The implication of this discussion for this study is important. ATE has been described as a multisite program that is being evaluated prospectively, but both the initial program model (Figure 3) and the current program model (Figure 1) were
developed after the program had been in operation for several years. While the initial program model depicted in Figure 3 may have been a closer representation of the normative program model, the current program model depicted in Figure 1 was clearly a descriptive model based on observed program operations. In addition, the data used for this study were aligned with the current program model thus responding to the question of why outcomes occurred was based on the program structure defined by the descriptive model.

Theory-Based Evaluation and Multisite Evaluation

Given that multisite evaluation is concerned with determining the conditions under which program succeeds or fails (Tushnet, 1995), theory-based evaluation is a natural compliment to multisite evaluation because program theory provides a framework for identifying the antecedents and outcomes around which data are collected (Rogers, et. al., 2000). Thus, evaluators can get closer to answering the question of why outcomes are occurring by having an apriori understanding of the underlying hypotheses about how a program works to achieve its intended outcomes (Rogers, et. al., 2000). The remainder of this discussion highlights several examples of theory-based evaluations as applied to the multisite context.

At the Federal level, the National Science Foundation (NSF) has provided multiple examples of where program theory can be associated with a multisite program, including the Instrumentation and Laboratory Improvement (ILI) program (NSF, 1998), the Instructional Materials Development (IMD) program (NSF, 2000), the Centers for Learning and Teaching (CLT) program, the Collaboratives for Excellence in Teacher Preparation programs (Lawrenz, et. al., 2003), as well as the ATE program (Hanssen, et. al., 2003; Lawrenz & Huffman, 2003).
These examples raise several issues regarding theory-based evaluation. First, each is a program, rather than an uncoordinated collection of responses to a problem, like Greenberg and colleagues described in their article on welfare to work programs (2003). This suggests that while multisite evaluations may target an issue, the application of theory-based evaluation in this context may be better suited to cases where a defined program exists.

While each program listed above had a general design, there is limited evidence to show that an explicit, normative program theory was in place prior to the start of the evaluation or prior to the start of the program. Nor was there evidence to show that the theory that was ultimately developed was used to guide the multisite evaluations of these programs. A possible exception to this was the CLT program, where evaluators developed a logic model for the program and then evaluation data collection methods were based on that model (Lawrenz & Huffman, 2003). However, since it is unclear when this logic model was developed (e.g., before or after the program began operating), experience tells us that it was probably developed after the program started, but before the evaluation began. Thus, Chen’s notion of normative program theory is rarely explicitly articulated in a multisite program context.

Second, the “theory” by which these example programs operate and against which they are evaluated varies in definition. For example, the theory related to the IMD program is an articulation of how materials are developed, disseminated, adopted, and implemented, and how this chain of events contributes to impact (see Figure 5). In this case, the theory that was used to guide the evaluation was developed through a review of past research on materials development and adoption.
practices (NSF, 2000), suggesting that this program theory was not keyed to the specific design of the IMD program. Rather, the program theory supported the evaluation design, which "was intended to inform NSF and others about the relationship of the development process to quality, and how both quality and approaches to marketing affect adoption and use. Further, the design provided contrasting information about marketing, adoption, and uses though identification of widely used non-NSF products and how they were disseminated, adopted, and used (NSF, 2000). As a result, this program theory was generic in part because it had to be appropriate for evaluating both IMD and non-IMD materials.

During the evaluation, the same types of information were collected from informants about IMD and non-IMD materials. This consistency in data collection enabled comparisons between the program-sponsored materials and the non-
program materials. Fundamentally, however, the program model allowed the evaluators to gather evidence and propose findings based on the underlying theory articulated by the program model.

Another variant on theory type is evident in the program logic model for ATE (Hanssen, et. al., 2003). The ATE program logic model used in this study was developed inductively after the program was in operation for nearly 10 years, thus it is a descriptive program model. This program logic model is based on the idea that more is better—e.g., if a project offers more professional development activities for current educators, then it will result in more improved programs. Thus, the causes in the ATE program are levels of intermediate outcomes and the effects are levels of subsequent outcomes.

This perspective differs somewhat from Lipsey (1993), who describes causes as events and effects as responses. Extending this thought process to ATE, one example manifestation of this program model might be that the adoption of the methods taught during a professional development activity would be the event, leading to the improved program, or response. In this case, however, the events and responses are based on measures of volume rather than on measures of actions. Clearly, this program model is descriptive in the sense that it is based on observed patterns of program execution and hypothesized relationships between different project activities.

Both examples of program theory described above, offer the opportunity for evaluators to provide answers to the question of why outcomes are produced. That, from the perspective of this study, is the critical value of theory-based evaluation.
Similarly, providing evidence about why outcomes are produced enables evaluators to provide defensible, formative feedback for programs.

**Structural Equation Modeling**

**Definition**

Structural equation modeling provides an excellent analytical tool for developing evidence to support formative evaluation feedback. Karl Jöreskog (2000) outlines the basic premise for using SEM:

Where ordinary regression methods no longer suffice [to answer the question of why results occur], and indeed give misleading results, is in purely observational studies in which all variables are subject to measurement error or uncontrolled variation and the purpose of the inquiry is to estimate relationships that account for variation among the variables in question [i.e., to explain why observed results occur]. This is the essential problem of data analysis in those fields where experimentation is impossible or impractical and mere empirical prediction is not the objective of the study. It is typical of almost all research in fields such as sociology, economics, [program evaluation], ecology, and even areas of physical science such as geology and meteorology. In these fields, the essential problem of data analysis is the estimation of structural relationships between quantitative observed variables. When the mathematical model that represents these relationships is linear we speak of a linear structural relationship. The various aspects of formulating, fitting, and testing such relationships we refer to as structural equation modeling (p. 1).

SEM compliments theory-based evaluation in that each element of the program logic model can be articulated as a latent construct. Each latent construct can be measured by observable indicators. The relationships between observable indicators and latent constructs are called factor models or measurement models (Jöreskog, 2000). The relationships between latent constructs represent the structural equation model (Jöreskog, 2000). The process of SEM estimates the strength of the relationships inherent in the measurement and structural models by
"fitting" the data to a specification of those relationships. Thus, SEM is based on the idea that (a) every theory implies a set of relationships, and (b) if the theory is valid, then the theory should be able to explain the patterns of relationships found in the available data (Kelloway, 1998).

The process for SEM typically involves 5 steps: (1) model specification, (2) identification, (3) estimation, (4) testing fit, and (5) re-specification (Bollen & Long, 1993). Model specification is akin to articulating a program theory or logic model. Identification addresses a critical statistical issue ensuring that the number of relationships (parameters) being estimated is less than the number of elements in the matrix of relationships available for analysis (Byrne, 1998). Estimation is the process of fitting the model and determining the best estimate for each parameter. Testing fit involves examining a variety of statistical measures to evaluate the degree to which a model fits (or does not fit) the available data. Finally, re-specification is the process of articulating competing program theories that may or may not fit the data better. These competing models are then typically estimated and their fit is evaluated (Jöreskog, 1993).

Although the number of different applications for SEM is too numerous to mention (Kelloway, 1998), Jöreskog (1993) differentiated between three general approaches to model testing: (1) strictly confirmatory, (2) alternative models, and (3) model generating. In a strictly confirmatory test, a researched defines a program theory and either accepts or rejects the model based on data gathered to test the model. In the alternative models case, a researcher specifies several models up front and selects one of the models based on an evaluation of model fit. The model generating case is most common. In this situation, a researcher specified
a tentative initial model and evaluates it to see if it fits the data. If the initial model did not fit, or does not provide an optimal fit, it can be modified and re-tested using the same data. In this process, several models are typically examined and the results may represent a stronger program theory and its associated structural equation model. Thus, the model generating case can best be described as a blending of the confirmatory and alternative models approaches.

In the context of theory-based evaluation, the strictly confirmatory approach to SEM is akin to evaluating a normative program model. Conversely, the model generating approach is analogous to development of a descriptive program model, which is based on actual observed program implementation. The alternative models approach bridges both the normative and descriptive program models.

In the context of this study, the model generating approach was appropriate given the starting point for the analysis is a descriptive program model. From this point, it was expected that the model would be refined and alternative models defined that might provide competing descriptions of how the ATE program operated.

SEM Examples in Evaluation

The literature is sparse in examples of multisite evaluations, or any evaluations, using SEM to establish relationships between program components. One can speculate that this is due to the emerging nature of both SEM (Kelloway, 1998) and program theory evaluation (Rogers, et. al., 2002). Regardless, Chen and Rossi (1983) suggested that SEM is a useful tool for analyzing linkages between treatments and outcomes. Smith (1990) also suggests that path analysis (an element of SEM) can be useful for developing and evaluating program theory. In addition, Hennessy and Greenberg (1999), Reynolds (1998), and Wang (1995)
provide three instructive examples of how this technique can be applied in evaluation settings.

**Integrating Program Theory and SEM**

Hennesy and Greenberg (1999) asserted that "the combination of programmatic theory and structural equation modeling can act as the basic intellectual machinery for designing and evaluating behavioral intentions" (p. 471). They argued that the critical starting point for applying these tools was a clearly articulated intervention theory that could be tested. Their view was that this theory would ideally be developed by the evaluator at the beginning of the program development process and then used throughout the program evaluation to guide program improvement. Thus, Hennesy and Greenberg are describing development of a normative program model and a prospective evaluation.

The context for their study was a randomized experiment to reduce sexual risk taking called the WINGS project. They demonstrated how the application of a program model and the testing of various paths within that model could lead to decisions about the most effective way to implement a program. They concluded that "SEM makes the programmatic theory accountable to program data because it combines data and theory in a way that is both intuitive and falsifiable" (p. 475).

In discussing why SEM and program theory evaluation have not been integrated more frequently, Hennesy and Greenberg cite (a) planning and management barriers, (b) data collection barriers, and (c) programmatic theory barriers. The primary planning barrier they argue is the requirement that to apply SEM, the evaluator must be brought in at the start of a program so that they can be involved in the articulation of a program theory. The data collection barrier, they
argue, is that the program theory developed may not be relevant in different program sites, thus the ability to collect consistent data across sites will be limited. Finally, the primary programmatic theory barrier is that the theory must be estimable, i.e., the number of parameters to be estimated must be less than the number of known quantities. Many program theories, then, are simply too complex because everything is thought to impact everything else.

Confirmatory Program Evaluation

Reynolds (1998) presents a different example based on child development intervention called the Child Parent Center Program. Reynolds termed this application "Confirmatory Program Evaluation," of which the primary focus was to determine the causal mechanisms within programs that produce program results. The data for this study were drawn from 360 students who attended school at six different program sites, thus the unit of analysis was the student and not the school. The program theory that guided this study held that a child's scholastic readiness was the result of (a) systematic language learning activities and (b) opportunities for family support experiences. The key measures for scholastic readiness were math and reading achievement test scores. This study used LISREL to determine the best fitting model from several alternatives.

Reynolds listed three primary limitations to confirmatory program evaluation. First, he indicated that this approach often requires more data collection than standard impact evaluation approaches. Second, he indicated that the findings are often dependent on the validity of the theory, i.e., if the theory is bad, then the findings can not be substantiated. Third, this method relies on the assumption that the program objectives can be accurately articulated, that the implementation of the
program can be accurately described, and that the causal mechanisms can be specified and measured.

Evaluating Program Implementation

Wang's (1995) study addresses a different problem. He was concerned with evaluating a model of program implementation that evaluators could use to study the implementation of site-managed educational programs. Like the IMD case discussed above, the research questions for the study focused on evaluating a theory of an implementation process, not a program theory that was guiding the program under study. Like the Reynolds case described above, data from over 200 educators at eight schools were used in the analysis; however the unit of analysis was the individual student, not the school building.

The model under study incorporated contextual variables, participant demographics, descriptions of the programs being implemented, and implementation results reflected by the degree of implementation and the quality of implementation. The exogenous variables in the model were those that described the participants in the study and the general environment in which the participants operated. The outcome variables were levels of program implementation. The underlying theory stated that the level of program implementation was a function of context, participant characteristics, and program descriptors.

The study effectively modeled the implementation process and offered two implications for program evaluation practice that are relevant to this study. First, Wang argued that evaluations of program implementation levels should address both the characteristics of the program and the implementation context. This implication implies a multisite context and varying implementation across sites. Second,
evaluators should identify closely correlated variables and eliminate redundancy in
the program model. This will save time and money in the evaluation process.

**Summary**

The examples presented above support the premises of this study by demonstrating that SEM can be applied in an evaluation context. These examples show how SEM is used to evaluate a structural model based that articulates the relationships between latent constructs. This structural model is closely related to an apriori specification of a program theory. This program theory is intended to reflect how the program is supposed to operate (normative) or actually operates (descriptive).

In the context of the ATE program, a descriptive program theory was developed by the evaluation team (see Figure 1). This descriptive program theory was articulated as a structural model (see Figure 2). This structural model was evaluated using SEM. The opportunity this represented was to clearly articulate the causal relationships between the defined elements of the structural model, which could then be related back to the program theory as specified by the evaluation team. Understanding these relationships would enable the evaluators to provide the program with formative feedback to improve the operation of the program with respect to increasing impact of the program through higher levels of outcomes.

**Relevance to This Study**

This study was concerned with the integration of three concepts as the basis for proposing an approach for providing defensible formative evaluation feedback, i.e., answering the question of why program outcomes occur. The general context
for this study was multisite evaluation. The specific context for this study was the ATE program. Program theory was a necessary element for answering the question of why; a descriptive program logic model was developed by the evaluation team and was used to guide the evaluation. Lastly, SEM is an analytical technique that can help explain the relationships between elements of the program theory. Because ATE was evaluated prospectively, data were available with regards to each of the key program constructs, to support the SEM process.
CHAPTER III

METHODOLOGY

The purpose of this chapter is to detail the methodology used for answering the research questions posed by this study. First, a detailed description of the study database is provided, including descriptions of the instrumentation, constructs, sample, and data preparation procedures. Included in this discussion are statements of nine hypotheses that are depicted in the ATE program logic model and that were defined by the structural model. Second, the analysis steps are detailed. The five-step structural equation modeling process described by Bolien and Long (1993) was used for this analysis. Included in this discussion are descriptions of the estimation and the model evaluation procedures. The section concludes with a description of the overall and comparative fit indices used to evaluate the structural models in this study.

By way of review, the research questions for this study were as follows:

In the context of multisite evaluation,

1. To what extent does the ATE logic model, in components or in its totality, fit the empirical data?

2a. If the model, in components or in its totality, can be fitted, what are the dynamics with which the programmatic characteristics contribute to program outcomes?

2b. If the empirical fit of the model is not optimal, can the current logic model be modified to improve the fit?
As described in Chapter 1, ATE is a multisite program that is being implemented in an uncontrolled manner. The evaluation of ATE is being conducted prospectively (Herrell & Sttraw, 2002). Figure 6 illustrates the ATE structural model that was evaluated in this study.

Figure 6. ATE program structural model

Database

The database for this study was the ATE Evaluation Database compiled by The Evaluation Center at Western Michigan University. This database contained data from the annual survey of ATE projects that was administered from 2000-2003. ATE projects that were active for at least one year at the time the survey was administered (usually between February and April) were invited to participate. Response rates for the survey approached 100 percent each year (Hanssen, et. al., 2003), thus, the responding projects represented the vast majority of all the projects funded by ATE.

Instrumentation

The survey instrument contained several hundred items (see Appendix D), a relatively small number of which were used in this study. Survey data were compiled
and stored in nine tables, one for each survey section. All projects that responded were asked to complete the first three survey sections: (1) basic information, (2) monitoring, and (3) the principal investigator overview. The six optional sections were (1) collaboration, (2) materials development, (3) professional development, (4) program improvement, secondary level, (5) program improvement, associate level and (6) program improvement, baccalaureate level. Projects were asked to complete these sections based on the program activities in which they engaged. For example, a project that collaborated with other institutions, development instructional materials, and developed a new curriculum at the associate level would be expected to complete optional sections 1, 2, and 5. For this study, it is important to note that each of the survey sections is represented by a latent variable in the ATE program structural model (see Figure 6).

Measures

The constructs pictured in Figure 6 were based on survey items from one or more survey sections. Table 3 provides an overview of each of the program constructs shown in the structural model. Each of the constructs and the measures for defining that construct are discussed below. Appendix E indicates each of the survey items for each observable measure and contains the detailed formulas for calculating each observed variable.

Project Characteristics

This construct represents the overall size and scope of the ATE project. Two indicators define this construct: the amount of the total current award amount that had been spent at the time of the survey (NSF) and the number of years the project had been operating (AGE). NSF was defined as the total award amount divided-by
the length of the award (e.g., 3 years) multiplied by the number of years the project had been in operation at the time it responded to the survey. AGE was defined as the difference between the survey date and the start date for the current award.

Table 3

**ATE program constructs**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Survey section(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Characteristics:</td>
<td></td>
</tr>
<tr>
<td>Size and scope of the project in terms of (1) the amount of support received from NSF at the time of the survey response and (2) the age of the project</td>
<td>Basic information</td>
</tr>
<tr>
<td>Organizational practices:</td>
<td></td>
</tr>
<tr>
<td>Rigor in operations as indicated by (1) use of evaluation and (2) participation in program monitoring</td>
<td>Monitoring</td>
</tr>
<tr>
<td>Collaboration:</td>
<td>Collaboration</td>
</tr>
<tr>
<td>Strength of linkages with other organizations as defined by (1) the amount of external support received and (2) the number of collaborating institutions</td>
<td></td>
</tr>
<tr>
<td>Materials Development:</td>
<td></td>
</tr>
<tr>
<td>Productivity in developing new educational materials defined by the numbers of (1) new courses, (2) modules, and (3) other materials developed</td>
<td>Materials development</td>
</tr>
<tr>
<td>Professional Development:</td>
<td></td>
</tr>
<tr>
<td>Productivity in developing educators defined by (1) the number of professional development opportunities offered and (2) the number of participants in those programs</td>
<td>Professional development</td>
</tr>
<tr>
<td>Program Improvement:</td>
<td></td>
</tr>
<tr>
<td>Productivity in producing new technician curricula as defined by (1) the number of programs offered, (2) the number of schools where programs are used, and (3) the number of courses contained in those programs</td>
<td>Program improvement</td>
</tr>
<tr>
<td>Student Impact:</td>
<td></td>
</tr>
<tr>
<td>Student impact defined by (1) the number of students served by a project, (2) the number of students who complete a specific program and (3) the number of students from that program who are placed in technician positions.</td>
<td>Program improvement</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Combined, these indicators define the overall scope of an individual project. This is integral to the program theory for several reasons. First, larger projects that receive larger awards should be expected to produce more results and impact more students than projects that received smaller awards. Second, projects that were in operation longer should have produced greater impacts.

Organizational Practices

This construct represented the degree to which a project applied sound principles of evaluation (EVAL) and monitoring (MNTR), to guide project work. EVAL was defined as the frequency with which a project used an evaluator. MNTR was defined as the number of different monitoring activities in which a project engaged. The underlying premise of this construct was that projects that do these things would have produced more outcomes than project that did not.

Collaboration

The collaboration construct described the degree to which projects were successful in securing external support (EXTSUP) for project activities and the number of different collaborating partners (PTNRS). EXTSUP was defined as the amount of monetary and in-kind support received by a project in the previous 12 month period. PTNRS was the total number of collaborators across four categories of organizations—(1) businesses, (2) educational institutions, (3) government entities, and (4) other types of institutions.

The collaboration measure was integral to the underlying ATE program model in that projects with high levels of collaboration were expected to make a larger impact on students through their project activities. The program model suggested that more collaboration contributed to (a) greater numbers of materials produced, (b)
higher demand for more professional development activities, and (c) more widespread adoption of curricula. Furthermore, collaboration was seen as being driven by project characteristics and organizational practices—larger projects and projects with sound practices could better leverage their resources to secure external funding and establish more external partnerships.

**Materials Development**

The first intermediate outcome within the ATE program model was materials development, which reflected productivity in this program area. Materials development was dependent on collaboration, project characteristics, and organizational practices—projects with higher funding levels, more rigorous practices, and more extensive collaborations should have been more productive in developing materials. Similarly, projects that produced more materials should have been positioned to make greater downstream impacts in professional development and program improvement.

Materials development was measured by determining a project’s productivity in developing three types of materials: courses (CRSE), modules (MDLE), and other types of materials (OTH_MAT). For each type of material, productivity was weighted based on the number of materials in various stages of development, from draft stage to commercial publication (see Appendix E). Similarly, course development was weighted more heavily than development of modules or other types of materials. The rationale for these weightings reflected the assumed level of effort needed to develop a particular type of material. For example, development of one course (a series of learning modules pieced together in a coherent instructional framework taking place over the course of an entire academic term) required greater
effort than development of one module (one learning event that may have appeared in multiple courses but that likely covers a narrow subject area and a shorter period of time).

Professional Development

This construct represented productivity in the area of professional development as defined by the total number of professional development opportunities (OPPORT) offered by a project in the preceding 12 months and the total number of participants (PARTS) who attended those opportunities. OPPORT was weighted depending on the type of activity offered—conferences, workshops, in-service, internships, online courses, and other types (see Appendix E). Weightings reflected the complexity of developing and delivering each type of activity. For example, developing an internship or in-service program was considered more complex than developing a 2-day workshop.

The position of this construct within the ATE program model reflected the relationship between professional development and collaboration, materials development, and program improvement. The program model suggested that higher levels of professional development productivity could be expected from projects with stronger collaborations and that developed more materials (thus requiring the need for more professional development).

Program Improvement

The program improvement construct was measured by the number of programs developed by a project (PRGMS), the number of courses contained in those programs (CRSES), and the number of schools (SCHLS) where the programs were offered. Numbers reported at each educational level (secondary, associate,
and baccalaureate) were aggregated for use in the analysis. These measures reflected the productivity of projects in developing new courses, integrating courses into curriculum programs, and promoting the adoption of these courses and programs in multiple schools.

Program improvement was positioned within the ATE program logic model such that results in this area were thought to be driven by the strengths of collaborations, productivity in materials development, and productivity in professional development. Conversely, program improvement was positioned as the sole antecedent to student impact. This portion of the model was based on the relationship defined by the ATE program model which indicated that direct student impact was achieved through students enrolling in ATE-supported courses and programs.

**Student Impact**

This construct was defined by three dimensions—the total number of students enrolled in at least one course (STDNTS), the number of students who completed a specific program (CMPLT), and the number of students who completed the program and were placed in technician related jobs (PLCMNT). STDNTS described the project's impact across all program-funded courses. CMPLT and PLCMNT were drawn from information about a single ATE-funded program selected by the responding project. These observable outcomes appear at the end of the program logic model and represent the ultimate outcomes for the ATE program.

**ATE Structural Equation Model**

The structural equation model that formed the basis for analysis in this dissertation is shown in Figure 7 below. This figure depicts both the structural model...
and the associated measurement models, including the 17 observed variables. As described above, project characteristics and organizational practices were hypothesized to be exogenous latent variables. Collaboration, materials development, professional development, program improvement, and impact were endogenous latent variables.

![Diagram of ATE program structural equation model](image)

**Figure 7. Full ATE program structural equation model**

The structural model is based on the descriptive program model developed by the evaluation team (Hanssen, et. al., 2003) and illustrates the relationships between latent variables in the ATE program structural model. The hypotheses listed below articulate the specific relationships between latent variables implied by the structural model. When the ATE structural model was estimated and evaluated, these hypotheses served as a guide for evaluating model fit.

**Hypothesis 1.** The size of a project will be positively related to the level of collaborations in which a project is engaged.

**Hypothesis 2.** The degree of rigor in a project’s organizational practices will be positively related to the level of collaborations in which a project is engaged.
Hypothesis 3. The level of collaborations will be positively related to the level of materials development productivity.

Hypothesis 4. The level of collaborations will be positively related to the level of professional development productivity.

Hypothesis 5. The level of collaborations will be positively related to the level of program improvement activity.

Hypothesis 6. The level of materials development productivity will be positively related to the level of professional development activity.

Hypothesis 7. The level of materials development productivity will be positively related to the level of program improvement activity.

Hypothesis 8. The level of professional development productivity will be positively related to the level of program improvement activity.

Hypothesis 9. The level of program improvement activity will be positively related to the level of student impact.

Sample

The sample used in this study was based on 168 survey responses from projects that responded to the ATE evaluation survey between 2000 and 2003 and that provided data for each of the constructs in the ATE program model. The second condition was satisfied if a project completed (a) the three required survey sections, (b) the collaboration, materials development, and professional development sections, and (c) at least one of the three program improvement sections.

The 168 survey responses were provided by 115 different projects. In cases where multiple responses were provided by a project, the responses were averaged to create a single response representing a "typical" program year. As a result, the
final sample used for analysis contained survey responses from 115 ATE projects. This sample size was at the lower range of acceptable practice given the number of variables in the analysis. For example, Bentler and Chou (1987) suggested there should be at least five subjects for each variable (though ideally one would have at least 10 subjects per variable). This ratio of subjects to variables was 115:17 or 6.8:1. While not ideal, this sample was large enough for analysis using structural equation modeling. It was expected, however, that the precision of standard error and parameter estimates would suffer given the small sample (Kelloway, 1998).

Data Preparation

Several steps were taken to prepare the data for analysis. First, raw data were coded and observed variables were calculated, if necessary, as described above and detailed in Appendix E. These tasks, were performed using SPSS software. This resulted in 17 variables for analysis with 115 cases. All variables were treated as continuous for analysis and the data did not contain any missing values. Table 4 provides descriptive statistics, which indicated that both the exogenous and endogenous variables contained sufficient variance to support the analysis. Variability of outcomes was critical for responding to the question of why.

Next, data were screened for univariate and multivariate normality using PRELIS (Jöreskog & Sörbom, 2003). Results revealed severe skewness and kurtosis for all but two observed indicators, EVAL and MNTR. Multivariate normality could not be supported, $X^2 (2, N=115) = 6354, p = .000$. With the exception of NSF and AGE, each observed variable demonstrated a floor effect such that most projects reported little (and in some cases no) activity for a given variable, while a
few projects exhibited large productivity for a given measure. This phenomenon was especially apparent in the program improvement and student impact measures.

Table 4

Descriptive statistics for ATE program measures (n=115)

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Mean</th>
<th>SD</th>
<th>Original Data Skew</th>
<th>Original Data Kurtosis</th>
<th>Normalized Data Skew</th>
<th>Normalized Data Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSF ($)</td>
<td>585,192</td>
<td>622,347</td>
<td>3.62</td>
<td>18.95</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>AGE</td>
<td>2.07</td>
<td>.78</td>
<td>1.22</td>
<td>4.61</td>
<td>0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>Organizational Practices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluator use (EVAL)</td>
<td>0.69</td>
<td>0.43</td>
<td>-0.82</td>
<td>-1.12</td>
<td>-0.73</td>
<td>-1.32</td>
</tr>
<tr>
<td>Monitoring (MNTR)</td>
<td>4.42</td>
<td>1.10</td>
<td>-0.43</td>
<td>-0.51</td>
<td>-0.11</td>
<td>-2.77</td>
</tr>
<tr>
<td>Collaboration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External support (EXT_SUP)</td>
<td>324,144</td>
<td>723,827</td>
<td>4.23</td>
<td>19.34</td>
<td>0.11</td>
<td>-0.34</td>
</tr>
<tr>
<td>Partnerships (PTNRS)</td>
<td>33.67</td>
<td>53.44</td>
<td>4.12</td>
<td>23.59</td>
<td>0.04</td>
<td>-0.15</td>
</tr>
<tr>
<td>Materials Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courses (CRSE)</td>
<td>65.93</td>
<td>118.60</td>
<td>3.19</td>
<td>11.54</td>
<td>0.23</td>
<td>-0.52</td>
</tr>
<tr>
<td>Modules (MDLE)</td>
<td>100.50</td>
<td>386.14</td>
<td>6.80</td>
<td>52.12</td>
<td>0.38</td>
<td>-0.62</td>
</tr>
<tr>
<td>Other materials (OTH_MAT)</td>
<td>29.98</td>
<td>242.15</td>
<td>10.23</td>
<td>107.78</td>
<td>1.43</td>
<td>0.82</td>
</tr>
<tr>
<td>Professional Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opportunities (OPPORT)</td>
<td>45.07</td>
<td>92.65</td>
<td>4.23</td>
<td>19.89</td>
<td>0.04</td>
<td>-0.15</td>
</tr>
<tr>
<td>Participants (PARTS)</td>
<td>149.05</td>
<td>244.74</td>
<td>3.16</td>
<td>11.44</td>
<td>0.03</td>
<td>-0.15</td>
</tr>
<tr>
<td>Program Improvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Programs (PRGMS)</td>
<td>7.10</td>
<td>13.36</td>
<td>3.41</td>
<td>12.01</td>
<td>0.20</td>
<td>-0.15</td>
</tr>
<tr>
<td>Schools (SCHLS)</td>
<td>12.11</td>
<td>23.79</td>
<td>5.16</td>
<td>34.48</td>
<td>0.11</td>
<td>-0.18</td>
</tr>
<tr>
<td>Courses (CRSES)</td>
<td>104.54</td>
<td>373.44</td>
<td>7.13</td>
<td>57.37</td>
<td>0.04</td>
<td>-0.15</td>
</tr>
<tr>
<td>Student Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. enrolled (STDNTS)</td>
<td>1,315</td>
<td>4,639</td>
<td>6.82</td>
<td>53.45</td>
<td>0.03</td>
<td>-0.12</td>
</tr>
<tr>
<td>No. completing (CMPLTS)</td>
<td>53.27</td>
<td>112.64</td>
<td>5.20</td>
<td>35.94</td>
<td>0.22</td>
<td>-0.51</td>
</tr>
<tr>
<td>No. placed (PLCMNTS)</td>
<td>31.11</td>
<td>86.53</td>
<td>5.42</td>
<td>35.60</td>
<td>0.44</td>
<td>-0.63</td>
</tr>
</tbody>
</table>

Note. Program constructs are referenced using Initial Capitals and italics. References to observable indicators use the abbreviations indicated above and in the previous text.

The literature supported accounting for non-normality by applying maximum likelihood (ML) or weighted least squares (WLS) estimation methods, but those solutions required large samples (Jöreskog, et.al., 2001), which were not available.

A suggested alternative was to normalize the data before analysis as long as the unit of measurement had no intrinsic meaning (Jöreskog, et.al, 2001). While the scales...
of the study variables were "meaningful," (e.g., dollars and numbers of students), this study was only interested in examining the relative contribution of variables to one another. Thus, the scales of individual measures were less important than the benefit to be gained by normalizing the data as suggested in the literature.

Based on these preliminary analyses, two procedures were applied to the variables prior to analysis. First, standard scores (i.e., z-scores) were calculated for each data element and a new data set was created based on these scores. Each variable in this data set had a mean of 0 and a standard deviation of 1. The added benefit of this procedure was to simplify the model estimation procedures by encouraging model convergence, which is more difficult when variable scales vary greatly. This procedure however, did not, nor was it expected to, correct the univariate and multivariate non-normality.

The second procedure normalized the standard score data set via PRELIS using a monotonic transformation (Jöreskog, et. al., 2002). This corrected the majority of the skewness and kurtosis for the observed variables. Exceptions were EVAL, $\chi^2 (2, N = 115) = 86.47, p = .000$; MDLE, $\chi^2 (2, N = 115) = 6.15, p = .046$; OTHMAT, $\chi^2 (2, N = 115) = 28.89, p = .000$; and PLCMNT $\chi^2 (2, N = 115) = 7.17, p = .028$. Multivariate normality was not achieved, $\chi^2 (2, N = 115) = 66.10, p = .000$, though the data appeared to be adequate for estimating the overall model fit and the significance of individual parameters.

Analysis

The analysis in this dissertation evaluated the structural model shown in Figure 7 above in an attempt to respond to the research questions posed for this
study. The structural model was derived from the ATE program logic model and the hypothesized relationships among the ATE program's primary constructs. That is, the model presented an initial response to answering the question of why results were occurring. The specific goal of the analysis, however, was to find one or more models that fit the data well statistically, taking all aspects of error into account, and had the properties of every parameter having a substantively meaningful interpretation.

The analytical engine used in this study was LISREL 8.54 (Jöreskog & Sörbom, 2003) and its companion program PRELIS. PRELIS was used to perform the data screening, data conversion and normalization, and computation of the covariance matrix and asymptotic covariance matrix used for model estimation. LISREL was used to estimate the measurement and structural models, produce fit indices, and support model modification through detailed reporting of the parameter characteristics.

The analysis strategy used in this study involved five steps: (1) specification, (2) identification, (3) estimation, (4) evaluation, and (5) re-specification (Jöreskog, 1993; Bollen & Long, 1993). For simplicity's sake, steps 4-5 were conducted simultaneously; these steps occurred in an iterative manner as the model was evaluated, re-specified, and re-evaluated until the best fitting model was determined.

**Specification**

The initial model specified above was based on the work of the ATE program evaluation team at Western Michigan University (Hanssen, et.al., 2003). As discussed in Chapter II, the ATE program model was a descriptive model in that it reflected the observed implementation of the ATE program after 10 years of...
operation. Similarly, the data that were used to estimate the model parameters were selected from an existing database that was limited to the outcomes-oriented evaluation questions contained in the annual ATE evaluation survey.

Identification

This step entailed estimating the measurement models for each construct to determine if it was possible to find unique values for the parameters of the specified model. In this study, each measurement model was "just identified," meaning that, the number of parameters being estimated was equal to the number of elements in the co-variance matrix (Byrne, 1998; Kelloway, 1998). As a result, the measurement models reproduced the covariance matrix perfectly, thus the test of such models was ultimately pointless (Stevens, 2002).

Estimation

This step involved selection of an estimation technique given the distributional properties of the variables being analyzed (Bollen & Long, 1993). There were two major elements to the estimation process: (1) selecting the correct matrix for analysis and (2) selecting an appropriate estimation method.

Matrix Selection. Most SEM theory was developed for application with the matrix of associations among the observed variables being a variance-covariance matrix (Thompson, 2002). Other common options under certain conditions include using a matrix of Pearson product-moment correlations or a matrix of polychoric correlations (Jöreskog & Sörbom, 2002). Generally, however, the preferred method is to analyze the variance-covariance matrix (Schumacker & Lomax, 1996) given that the use of a correlation matrix can lead to imprecise estimates of the standard errors for the parameter estimates (Boomsma, 1982).
Additionally, when continuous data are not multivariate normal, and the sample size is small, then the calculation of an asymptotic covariance matrix is recommended (Byrne, 1998; Jöreskog & Sörbom, 2003). This matrix is derived from the covariance matrix based on the raw data and is used to approximate the population covariance matrix, yielding a more accurate parameter estimates and fit indices.

For this study, both the variance-covariance matrix and asymptotic covariance matrix, for the normalized, standard-score data, were analyzed. However, the covariance matrix of standard scores was found to be equivalent to the correlation matrix of the original data. In essence, then, this study analyzed the correlation matrix, though it was necessary to "trick" the LISREL program into thinking it was analyzing a covariance matrix in order to enable application of the recommended estimation method. While not ideal, Kelloway (1998) argued that if one is only concerned with the pattern of relationships among the variables, analyzing a correlation matrix is an appropriate choice. Additionally, use of the correlation matrix can result in more conservative estimates of parameter significance, which is generally held to be desirable in statistics.

**Estimation Method.** The estimation process involved selecting a fitting function to minimize the difference between the population and sample covariance matrices (Schumacker & Lomax, 1996). LISREL offered a choice of seven estimation methods: instrumental variables method (IV), two-stage least squares (TSLs), unweighted least squares (ULS), generalized least squares (GLS), maximum likelihood (ML), weighted least squares (WLS), and diagonally weighted least squares (DWLS) (Byrne, 1998). Each of these methods has unique purposes.
and underlying assumptions based on the distributional properties of the data. IV and TSLS are not based on iterative processes and are often used to provide fast, initial parameter estimates to aid in model specification. The other methods compute estimates iteratively using starting values provided by IV or TSLS (starting values produced by IV are used for ULS. TSLS provides values for the other estimation methods) (Byrne, 1998).

ML and GLS are the most common estimation methods (Kelloway, 1998) and LISREL defaults to using ML unless another method is specified. For ML estimation if the sample is large and the observed variables are multivariate normal, then this is the recommended choice; GLS should be used if the variables are not multivariate normal (Kelloway, 1998). Neither of these approaches was appropriate for this study given the small sample size.

Browne (1987), however, described an estimation process called Robust Maximum Likelihood (RML), which is used when the data to be analyzed (a) are continuous, (b) do not follow a multivariate normal distribution, and (c) the number of cases is not large. RML requires an estimate of the asymptotic covariance matrix of the sample variances and covariances (Jöreskog & Sörbom, 2003). It is implemented in LISREL by providing both matrices for analysis and specifying the ML method of estimation (Mels, 2003). The requirements for using RML matched the conditions of this study and were used to fit and evaluate the model.

Evaluation and Modification

The model specified above was evaluated in three stages: (1) overall model fit was assessed, (2) detailed model fit was assessed, and, after suitable alternative models were identified, (3) competing models were compared.
Overall model fit. Three fit indices were used to assess overall model fit. Each is described briefly below.

Chi-square. The Chi-square statistic is used to test the hypothesis that the original population matrix is equal to the matrix reproduced from the test model. The preferred result is to not reject this hypothesis, where rejecting the null hypothesis indicates that the sample and population matrices are not equivalent (Stevens, 2002). Satorra and Bentler (1994) described an adjustment to the chi-square fit index for use in cases where the sample size is small and data are non-normal. Given that both of these conditions were present in the study data, the Satorra-Bentler Chi-square was used as the basic measure of overall fit. LISREL generates this statistic automatically when using ML estimation and providing the asymptotic covariance matrix for analysis (Jöreskog & Sörbom, 2003).

Root mean square error of approximation (RMSEA). The lack of fit of the hypothesized model to the population is known as the error of approximation (Stevens, 2002) and the RMSEA (Steiger, 1990), is a standardized measure of this error. MacCallum (1995) called the RMSEA essentially a measure of lack of fit per degree of freedom. RMSEA values of .05 or less indicate a close model fit and values up to .08 suggest a reasonable fit (Brown & Cudeck, 1993). The RMSEA is also reported with a 90% confidence interval, which aided in interpretation of these results. These benchmarks guided the evaluation of overall model fit based on this index.

Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI). The GFI and AGFI were used as measures of absolute fit because they compare the hypothesized model with the null model (Byrne, 1998). The null model is one where
there are no relationships between the observed variables, i.e., all of the values in
the variance-covariance matrix are 0. The AGFI differs from the GFI in that it adjusts
for the number of degrees of freedom in the model being estimated. Thus, there is a
penalty applied to these indices for incorporating additional parameters, which
typically results in stronger overall fit, into the model (Byrne, 1998). Values for the
GFI and AGFI range from 0 to 1, with values between .9 and 1.0 indicating good fit.

**Detailed Model Fit.** The next stage in model evaluation involved evaluation of
the fit of individual paths (i.e., evaluation of hypothesized relationships). The tools
used for evaluating detailed model fit were t-tests, standardized residuals, and
modification indices. Each of these statistics was used to locate the source of
misspecification and to suggest how the model should be modified to fit the data
better. Any misspecification was considered, but the ultimate decision to re-specify
the model was guided by the underlying program logic model.

**T-tests.** This test was conducted for each parameter estimate to determine if
the estimate was statistically different from zero (Byrne, 1998). The t-statistic was
calculated by dividing the parameter estimate by its standard error. Based on an
alpha level of .05, this statistic needed to be greater than +/- 1.96 before the
parameter could be considered significant. For this study, statistical significance was
reported at alpha levels of .10, .05, and .01, indicating t values of +/- 1.645, +/- 1.96,
and +/- 2.576, respectively.

**Standardized Residuals.** A residual is an observed minus a fitted covariance
(variance). A standardized residual is a residual divided by its estimated standard
error. These residuals exist for every pair of observed variables. Fitted residuals
depend on the unit of measurement of the observed variables, thus if the variances
of the variables are considerably different from one another, then it is difficult to know whether a fitted residual should be considered large or small. Standardized residuals are independent of the units of measurement of the variables and thus provide a "statistical" metric for judging the size of a residual.

A large positive residual indicated that the model underestimated the covariance between the two variables, suggesting that a path should be added to better account for the covariance between the variables. Conversely, a large negative residual indicated that the model overestimated the covariance between the variables, suggesting a path should be deleted.

The standardized residuals were examined collectively in two plots: (1) a stem-leaf plot and (2) a Q-plot. A good model was characterized by a stem-leaf plot where the residuals were symmetrical around zero, with most in the middle and fewer in the tails. An excess of residuals on the positive or negative side indicated that residuals may have been systematically under- or overestimated. In the Q-plot, a good model was characterized by points on a scatter plot of standardized residuals (x-axis) and normal quantiles (y-axis) falling approximately on a 45 degree line. Deviations from this pattern indicated specification errors in the model, non-normality in the variables or nonlinear relationships among the variables. In particular, standardized residuals that appeared as outliers in the Q-plot were indicative of a specification error in the model.

Modification indices. A modification index (Sörbom, 1989) was computed for each fixed or constrained parameter in the model. Each modification index measured how much the chi-square statistic was expected to decrease if the particular parameter was set free (to be estimated) and the model re-estimated.
Thus, the modification index was approximately equal to the difference in chi-square between two models in which one parameter is fixed or constrained and free in the other, all other parameters being estimated in both models.

Modification indices were used in the process of model evaluation and modification in the following way (Jöreskog & Sörbom, 2003). If chi-square was large relative to the degrees of freedom, the modification indices were examined and the parameter with the largest modification index was set free, as long as freeing that parameter could be meaningfully interpreted within the context of the program model being evaluated. If it did not make sense to relax the parameter with the largest modification index, the second largest modification index was considered, etc., but only one model change was made before the model was re-estimated.

This process was repeated for each model estimated and, if this step led to a modified model, repeated on each modified model.

Model Comparison. Competing models, to the degree that they were empirically and substantively justified, were compared to determine which program logic model represented the best fit for the sample data. Models were compared using (a) the Satorra-Bentler scaled chi-square difference test (Satorra & Bentler, 2001) and (b) the comparative fit indices described below. The Satorra-Bentler scaled chi-square difference test was used, because the normal theory chi-square difference test was found to not yield the correct scaled difference test statistic when applied to two Satorra-Bentler scaled chi-square test statistics (Satorra, 2000). However, the interpretation of the Satorra-Bentler scaled chi-square difference test was identical to that for the normal theory chi-square difference test. If the scaled chi-square difference test statistic, given the difference in degrees of freedom, was
significant, the models were interpreted to be different. Each of the additional comparative fit indices is described below.

**Normed Fit Index (NFI) and Non-normed Fit Index (NNFI).** The NFI represents the increment in fit obtained by using the hypothesized model relative to the fit of the null model. Values range from zero to one, with higher values (e.g., values of .90 or greater) indicating greater improvement in fit (Stevens, 2002). For example, an NFI value of .8 means that the model is 80% better fitting than the null model (Kelloway, 1996). The NNFI applies a similar logic, but adjusts the NFI for degrees of freedom.

**Akaike Information Criteria (AIC).** The AIC measure was used to compare models with differing numbers of latent variables (Akaike, 1987). Thus, like the GFI and AGFI in assessing overall model fit, parsimony was taken into account by indicating both model fit and whether or not a model was over identified (Schumacker & Lomax, 1998). Smaller AIC values were preferred.

**Summary**

The methodology used in this study was appropriate based on the problem identified and the rationale for this study. The following summarizes the critical steps and decisions used in this study.

1. **Instrumentation.** The data were obtained from the ATE evaluation database that was compiled through administration of the annual ATE program evaluation survey from 2000-2003.
2. **Measures.** Seventeen (17) variables were created based on survey items from the annual ATE evaluation survey. These variables were related to seven latent constructs that formed the basis of the ATE structural equation model.

3. **Sample.** ATE projects that responded to the survey and completed all nine survey sections were selected for this study. In cases where a project responded to the survey more than once from 2000-2003, the average survey response was used for analysis. The sample size was 115.

4. **Data preparation.** The data were converted to standard scores to reduce the impact of widely varying measurement scales. The standard score data were normalized to correct for multivariate non-normality.

5. **Estimation.** The covariance and asymptotic covariance matrices were analyzed using robust maximum likelihood estimation. This method was supported for small samples of continuous, non-multivariate normal data.

6. **Model evaluation.** Overall model fit was evaluated using the Satorra-Bentler Chi-Square, the RMSEA, and the GFI and AGFI.

7. **Re-specification.** Detailed model fit was evaluated using t-tests, standardized residual plots, and modification indices. The model was re-specified and re-estimated based on these indicators, as long as the suggested changes were meaningful given the context of the ATE program model.

8. **Model comparison.** Competing models were compared using the Chi-square difference test, the NFI and NNFI, and the AIC.
CHAPTER IV

RESULTS

This study used the ATE program evaluation as the context for answering the following questions:

In the context of multisite evaluation,

1. To what extent does the ATE logic model, in components or in its totality, fit the empirical data?

2a. If the model, in components or in its totality, can be fitted, what are the dynamics with which the programmatic characteristics contribute to program outcomes?

2b. If the empirical fit of the model is not optimal, can the current logic model be modified to improve the fit?

The results of this study are presented as follows. First, detailed structural equation modeling results based on the full ATE program model are presented and discussed. The discussion of these results addresses the seven hypotheses that pertain to the relationships between the latent variables (program components) and addresses research questions 1 and 2a. Second, three alternative models are defined and discussed. A comparison of fit between the full program model and the alternative models is used to address question 2b. Finally, summary responses to the stated research questions are provided based on these analyses.
Overall Model Fit

The full program model as depicted in Figure 7 was analyzed using the covariance and asymptotic covariance matrices with robust maximum likelihood (RML) estimation as implemented in LISREL 8.54 (Jöreskog & Sörbom, 2003). Based on this analysis, the structural model provided a marginal fit to the data, $\chi^2$ $(109, N = 115) = 169.53, p < .001; RMSEA = .07; GFI = .84 AGFI = .77.$

Because the model fit was marginal, a means was sought to improve model fit without removing any parameters from the model. Modification indices showed that allowing the path from Project Characteristics to MNTR to be freely estimated would result in a significant improvement in model fit ($MI = 108.7; EC = 20.36$). This change, however, was rejected on substantive grounds. The next largest modification index showed that allowing the covariance between the error terms of CMPLT and PLCMNT to be freely estimated, would result in a substantial improvement in model fit ($MI = 26.78; EC = .39$). This suggested modification was defensible in that it was reasonable to expect projects with large numbers of students completing their programs to also place larger numbers of students in technician positions (Hanssen, et. al., 2003). Thus, a stronger fit was sought by allowing the covariance between error terms of CMPLT and PLCMNT to be freely estimated. The re-estimated model provided an acceptable fit to the data, $\chi^2$ $(108, N = 115) = 140.98, p = .0018; RMSEA = .05; GFI = .86 AGFI = .81$, suggesting that the data adequately represent the ATE program model, as conceptualized in the ATE program model shown in Figure 7.
Detailed Model Fit

Given an acceptable overall model fit, detailed model fit was evaluated as follows: (a) the standardized path coefficients were evaluated for consistency with the program theory and statistical significance; (b) standardized residuals were examined; (c) the measurement models were evaluated for statistical significance; and (d) modification indices were examined to determine if specific parameters should be fixed or relaxed. The goal of these analyses was to determine what alternative models might be specified to provide a better representation of the program logic.

Structural Path Coefficients. Figure 8 shows the standardized path coefficients for the structural model. Although the overall model fit was adequate, these results were mixed in terms of consistency with the stated program logic model. First, Hypothesis 1 was not supported. The path coefficient of .05 was not significant and indicates that no relationship existed between Project Characteristics and Collaboration. Second, Hypothesis 2 was also not supported. Though statistically significant, the path coefficient of -.65 indicated that the relationship between Organizational Practices was inversely related to Collaboration. Though this relationship was statistically significant, the high standard error (.39) also suggests this coefficient would probably not remain stable in another sample. Combined, these findings suggest that the exogenous portion of the model is not supported by the available data.

Hypotheses 3, 4, and 5 were supported by the results. The path coefficients describing the relationships between Collaboration and Materials Development (1.59), Collaboration and Professional Development (1.11), and Collaboration and
Program Improvement (.91), hypotheses 3, 4, and 5, respectively, were consistent with the stated program logic model, supporting the claim that higher levels of Collaboration led to higher productivity in these three areas.

![Diagram]

Note. Standard errors are reported in parentheses.

Figure 8. ATE program model with standardized path coefficients

Hypothesis 6, the relationship between Materials Development and Professional Development, was not supported. The path coefficient of -.26 showed a weak, inverse relationship that contradicted the stated program logic. Hypothesis 7 was also not supported; the path coefficient of -.27 showed that a weak inverse relationship between Materials Development and Program Improvement existed. Additionally, neither of the path coefficients for hypothesis 6 or 7 was statistically significant.

Combined, these findings suggested that the latent variable, Materials Development, may be poorly defined (although the measurement model for Materials Development was supported by the analysis). An alternative explanation, grounded
in the program operations, was also possible. In the last few years, the ATE program has de-emphasized materials development as a tangible outcome, preferring to imbed these outcomes in other activities, such as program improvement (Hanssen, et. al., 2003). In other words, materials development that is subsumed within program improvement (or professional development) may not be accurately captured by the survey items underlying this construct. This may have contributed to non-statistically significant relationships between these constructs.

Hypothesis 8, which described the direct relationship between Professional Development and Program Improvement, was supported by the analysis, although the path coefficient of .36 was not statistically significant.

Finally, Hypothesis 9 was supported. The path coefficient was .95 and was statistically significant, demonstrating a strong direct relationship between Program Improvement and Student Impact. This indicated that high levels of program improvement led to high levels of student impact.

**Measurement Model Evaluation.** Table 5 displays the factor loadings for the measurement models underlying the ATE program structural model. Based on these results, there were four primary weaknesses detected in the measurement model. First, the measurement of Project Characteristics was not informative as shown by the weak and non-significant factor loadings. Second, the measurement of Organizational Practices was marginally supported, though the factor loading for monitoring was moderately strong and statistically significant. Third, the measurement of Materials Development, though statistically significant, had weak factor loadings. Specifically, the loading of OTH_MAT was particularly weak and statistically significant only at the .10 level. Overall, this weak measurement could
have been a contributing factor for the negative path coefficients from Materials Development to both Professional Development (hypothesis 6) and Program Improvement (hypothesis 7), and for the highly positive coefficient from Collaboration to Materials development (hypothesis 3). Fourth, there was evidence to suggest that CMPLTS and PLCMNTS were redundant measures because of the influence that the correlation of the error terms for these measures ($r = .39, p < .01$) had on the overall model fit.

Table 5

Standardized parameter estimates for measurement models (full ATE program model)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVAL</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNTR</td>
<td>-.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXT_SUP</td>
<td></td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTNRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.72***</td>
</tr>
<tr>
<td>CRSE</td>
<td></td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDLE</td>
<td></td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td>.11**</td>
</tr>
<tr>
<td>OTH_MAT</td>
<td></td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td>.10*</td>
</tr>
<tr>
<td>OPPORT</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td>.81***</td>
</tr>
<tr>
<td>PARTS</td>
<td></td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td>.81***</td>
</tr>
<tr>
<td>PRGMS</td>
<td></td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td>.78***</td>
</tr>
<tr>
<td>SCHLS</td>
<td></td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td>.71***</td>
</tr>
<tr>
<td>CRSES</td>
<td></td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
<td>.83***</td>
</tr>
<tr>
<td>STDNTS</td>
<td></td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
<td>.86*</td>
</tr>
<tr>
<td>CMPLTS</td>
<td></td>
<td>.43</td>
<td></td>
<td></td>
<td></td>
<td>.43***</td>
</tr>
<tr>
<td>PLCMNTS</td>
<td></td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td>.46***</td>
</tr>
</tbody>
</table>

*Statistical significance of the first observable variable for each endogenous latent variable was not evaluated as these values are fixed in order to appropriately scale the associated latent variable (Kelloway, 1998). ***$p < .01$, **$p < .05$, *$p < .10$.
Residual Analysis

Analysis of standardized residuals was conducted (1) to provide an additional perspective on the overall model fit and (2) to identify potential sources of misspecification that could be corrected within any alternative models. A stem-leaf plot of the standardized residuals (see Figure 9) was relatively evenly distributed around zero, with only a slightly larger number of negative residuals. This suggested that the model was slightly overestimated (Jöreskog, 2003). This finding was not surprising given the ratio of parameters estimated (44) to the sample size (115). Overall, however, the stem-leaf plot supported the conclusion that the overall model fit the data.

Figure 9. Stem-Leaf plot of standardized residuals

A Q-plot of the residuals was consistent with the stem-leaf plot and revealed the presence of several large residuals (shown in Figure 9 as those residuals with values +/- 1.96) indicating sources of misspecification. Large negative residuals were found between PLCMNTS, and PRGMS and CRSES, and between OTH_MAT and STDNTS. Large positive residuals were found between OTH_MAT, and NSF.
and PARTS. These findings were consistent with analysis of the measurement model, which suggested that these observed variables presented problems for properly fitting the model to the data.

**Modification Indices**

Analysis of the modification indices indicated that the overall model fit could be further improved by allowing additional parameters to be freely estimated, although, additional paths were not supported between any of the latent variables in the model. The largest modification index supported allowing CMPLTS to load on Collaboration ($MI = 24.43$, expected change = -4.71). This change, however, could not be supported within the context of the ATE program model in that there was insufficient evidence available to show that the number of students who complete an ATE funded program was an indicator of collaboration. Other modification indices were not significantly large to warrant further model re-specification.

**Conclusions**

The assessment of detailed model fit suggested that alternative models may present a better fit for the sample data. Findings indicated that both Project Characteristics and Organizational Practices could be eliminated from the overall program model—there was insufficient evidence to support the claim that these constructs were related to Collaboration. This change placed Collaboration as the sole exogenous factor in an alternative ATE program model. Similarly, the elimination of OTH_MAT as an indicator of Materials Development and PLCMNTS as an indicator of Student Impact simplified the measurement of these factors enabling a more streamlined evaluation of their contribution to the program model.
Comparative ATE Program Models

Based on the findings above, three alternative models were specified and estimated. These models hypothesized similar relationship to those articulated for the full ATE program model. The key differences between the alternative models and the full model were the elimination of selected latent and observed variables. The following describes each model and presents the estimation results.

**Collaboration-Impact (CI) Model**

The CI model (see Figure 10), excluded *Project Characteristics* and *Organizational Practices* due to the poor measurement of these constructs. In addition, OTH_MAT was eliminated as an indicator of *Materials Development* due to the weak factor loading on this variable, and PLCMNTS was eliminated as an indicator for *Student Impact* based on the apparent redundancy of this variable with CMPLTS.

![Figure 10. CI model with standardized path coefficients](image-url)

Note. Standard errors are reported in parentheses.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
This model provided an overall strong fit to the data, $X^2 (37, N = 115) = 41.95, p = .26; RMSEA = .034; GFI = .93 AGFI = .87). While not all of the path coefficients were statistically significant, they were of sufficient magnitude and direction to support the revised program logic model.

The exception to this finding was the path between professional development and program improvement, which were inversely related. This finding suggests that high productivity in professional development is moderately related to lower productivity in program improvement. From a substantive perspective, this finding can be explained by reasoning that, given limited resources, allocation of time and money to professional development would necessarily detract from program improvement activities. This suggests projects that elected to engage in both activities were only able to perform one or the other satisfactorily. However, this conclusion cannot be supported by the existing program logic model (Hanssen, et al, 2003; NSF, 2003c).

An alternative explanation is that the strong relationship between Materials Development and Program Improvement created a suppression effect, thus creating a negative relationship between Professional Development and Program Improvement, which in turn balanced the model. This explanation seemed plausible given the weak measurement of Materials Development may have resulted in an artificially large path coefficient between Materials Development and Program Improvement.

**Collaboration-Impact No Materials Development (CI no MD) Model**

Based on the finding that (a) the measurement of Materials Development was poor and (b) that accounting for this productivity may be imbedded in Professional Development.
Development and Program Improvement activity, a second alternative model eliminated the Materials Development construct from the CI model (see Figure 11). This model was fully nested within the CI model presented above.

This model fit the data extremely well, $\chi^2 (23, N = 115) = 24.57, p = .37$; RMSEA = .024; GFI = .95 AGFI = .89, and supported the overarching ATE program logic model. Path coefficients relating Collaboration and Professional Development, and Collaboration and Program Improvement were strong and statistically significant. The path from Professional Development to Program Improvement was positive (in direct contraction to the CI model), however, the coefficient was small and it was non-significant. This finding also supports the suppressor hypothesis which argued that Materials Development created a suppression effect in the CI model. With the removal of Materials Development, the path coefficient between Professional Development and Program Improvement reverted to a level that supported the ATE program logic.

Figure 11. CI no MD model with standardized path coefficients

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Full Model No Materials Development (FM no MD) Model

The third comparative model was the full ATE program model, excluding the materials development component (see Figure 12). This model provided a very strong fit to the data, $X^2 (69, N = 115) = 74.64, p = .30; \text{RMSEA} = .027; \text{GFI} = .91 \text{AGFI} = .86$.

![Diagram of model](attachment:figure12.png)

**Note.** Standard errors are reported in parentheses.

**Figure 12.** FM no MD model with standardized path coefficients

The path coefficients for this model were consistent with the previous analyses. First, the *Project Characteristics* and *Organizational Practices* path coefficients mirrored those from the full program model. This supported the finding that these constructs were not related to *Collaboration*. Second, the path coefficients between *Collaboration* and *Professional Development* and *Collaboration* and *Program Improvement* were nearly identical to those from the CI no MD model, as was the coefficient from *Professional Development* to *Program Improvement*. These findings supported the assertion that poor measurement of the *Materials Development* construct adversely impacted the overall model fit and the strength of the latent variable path coefficients.
**Comparative Fit**

Table 6 displays the overall fit indices and comparative fit indices for each model estimated in this study. These results indicate that the CI no MD model is the strongest fitting model based on evaluation of the NFI, NNFI, and AIC.

### Table 6

**Comparative model fit**

<table>
<thead>
<tr>
<th>Model</th>
<th>$X^2$</th>
<th>df</th>
<th>RMSEA w/ CI</th>
<th>GFI</th>
<th>AGFI</th>
<th>NFI</th>
<th>NNFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model</td>
<td>140.98*</td>
<td>108</td>
<td>.052 (.023, .074)</td>
<td>.86</td>
<td>.81</td>
<td>.87</td>
<td>.93</td>
<td>230.98</td>
</tr>
<tr>
<td>FM no MD</td>
<td>74.64</td>
<td>69</td>
<td>.027 (.000, .063)</td>
<td>.91</td>
<td>.86</td>
<td>.92</td>
<td>.98</td>
<td>146.64</td>
</tr>
<tr>
<td>CI</td>
<td>41.95</td>
<td>37</td>
<td>.034 (.000, .077)</td>
<td>.93</td>
<td>.87</td>
<td>.93</td>
<td>.97</td>
<td>99.95</td>
</tr>
<tr>
<td>CI no MD</td>
<td>24.57</td>
<td>23</td>
<td>.024 (.000, .082)</td>
<td>.95</td>
<td>.89</td>
<td>.96</td>
<td>.98</td>
<td>68.57</td>
</tr>
</tbody>
</table>

*p < .01

Table 7 displays the results of the Satorra-Bentler scaled chi-square difference tests for each model comparison.

### Table 7

**Satorra-Bentler scaled chi-square difference tests**

<table>
<thead>
<tr>
<th>Models compared</th>
<th>$X^2$</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model, FM no MD</td>
<td>65.84**</td>
<td>39</td>
</tr>
<tr>
<td>Full model, CI</td>
<td>99.78*</td>
<td>71</td>
</tr>
<tr>
<td>Full model, CI no MD</td>
<td>117.19**</td>
<td>85</td>
</tr>
<tr>
<td>FM no MD, CI</td>
<td>32.29</td>
<td>32</td>
</tr>
<tr>
<td>FM no MD, CI no MD</td>
<td>50.13</td>
<td>46</td>
</tr>
<tr>
<td>CI, CI no MD</td>
<td>17.52</td>
<td>14</td>
</tr>
</tbody>
</table>

**p < .01, *p < .05**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
These results indicate that each nested model represents a stronger fit to the data than the full model, and that the nested models are not statistically different from each other. These findings suggest that the measurement of Materials Development was poor and was reflected in the model estimation results. Similarly, the measurement of Project Characteristics and Organizational Practices did not make significant contributions to the model estimation such that the overall fit was not substantively different, with or without these factors in place. In the interests of parsimony, therefore, it was reasonable to declare that the Cl no MD represented the best fit of the data. This model can be described as follows:

1. Higher levels of Collaboration contribute to higher productivity in Professional Development.
2. Higher levels of Collaboration contribute to higher levels of Program Improvement.
3. Higher levels of Professional Development may be related to higher levels of Program Improvement.
4. Higher levels of Program Improvement lead to higher levels of Student Impact.

Summary

The results of this study were keyed to the three research questions. Each question is addressed below.

**Fit of the Full ATE Logic Model**

As demonstrated above, the full ATE program model fit the data reasonably well. Overall fit was reasonable after a minor re-specification, which allowed the covariance between the error terms for CMPLTS and PLCMNTS to be freely
estimated. Standardized path coefficients were consistent in direction, with the exception of the path between Organizational Practices and Collaboration. Although the overall fit obtained was reasonable, not all of the path coefficients were statistically significant, indicating potential specification problems within the model.

A detailed fit assessment revealed several potential sources of misspecification within the measurement models for Project Characteristics, Organizational Practices, Materials Development, and Student Impact.

Program Characteristics and Outcomes

Several of the hypotheses regarding the dynamics between program characteristics and outcomes were supported, while several others were not. First, there was support for the hypothesis that levels of Program Improvement were directly related to levels of Student Impact. Second, Collaboration was positively related to Materials Development productivity, Professional Development, and Program Improvement. Third, Professional Development was positively related to Program Improvement. Again, while not all of these path coefficients were statistically significant, the magnitude and direction of the coefficients provided at least some support for the hypotheses.

Hypotheses that were not supported were as follows: (1) Materials Development productivity was directly related to Professional Development productivity and Program Improvement levels; (2) Project Characteristics was directly related to Collaboration; (3) Organizational Practices were directly related to Collaboration. Combined, these findings suggested the need for some re-articulation of the program logic model and improvement in the measurement of the observed variables.
Alternative Models

Three alternative models were defined to determine if the data provided a better fit to a different articulation of the program logic. These alternative models were based on two significant changes: (1) the removal of Project Characteristics and Organizational Practices as exogenous latent variables, resulting in Collaboration as the exogenous factor in the program logic model, and (2) removal of Materials Development. Additionally, the measurement models for Materials Development and Student Impact were refined based on the detailed assessment of model fit conducted for the full ATE program model.

Each of the three alternative models provided a strong fit for the data that was statistically better than the full model fit. The full model fit worse than the full model without materials development; it also fit worse than two alternative models that used Collaboration as the exogenous variable—one with and one without Materials Development as a factor.

When parsimony was considered, it appeared that the most appropriate model for the data was the CI model without Materials Development. This model presented a much simplified view of the ATE program and indicated that a project’s level of collaboration was the critical factor in driving professional development productivity and the level of program improvement. Of course, levels of Program Improvement were directly related to Student Impact, which was the critical finding that was consistently supported throughout the analysis.
CHAPTER V
DISCUSSION

Summary

This study was concerned with formative evaluation and defining an approach for answering the question of why program results occurred. To accomplish this, the study integrated multisite evaluation, theory-based evaluation, and structural equation modeling. The Advanced Technological Education program provided the specific context for this study. ATE was a large multisite program and the program evaluation of ATE was conducted by The Evaluation Center at Western Michigan University. The central component of this evaluation was an annual survey of funded projects. The data from this survey provided the basis for answering the following research questions:

In the context of multisite evaluation,

1. To what extent does the ATE program logic model, in components or in its totality, fit the empirical data?

2a. If the model, in components or in its totality, can be fitted, what are the dynamics with which the programmatic characteristics contribute to program outcomes?

2b. If the empirical fit of the model is not optimal, can the current logic model be modified to improve the fit?
In order to respond to these questions, a structural model was developed (see Figure 7) based on a logic model that depicted how the ATE program operated (see Figure 1). The structural model contained seven latent constructs and nine hypotheses were developed to articulate the expected relationships between these constructs. Seventeen (17) observed variables formed the basis for measuring the latent constructs. Each variable was constructed from one or more evaluation survey items (see Appendix E). The data were standardized, so that each variable had a mean of zero and a standard deviation of one, and normalized using a monotonic transformation to correct for multivariate non-normality. Robust maximum likelihood estimation of the covariance and asymptotic covariance matrices was used to estimate overall model fit.

In response to the first research question, the overall fit was reasonable after a minor re-specification of the model, which allowed the covariance between the error terms for CMPLTS and PLCMNTS to be freely estimated. Standardized path coefficients were consistent in direction with the stated hypotheses, with the exception of the path between Organizational Practices and Collaboration. Although the overall fit obtained was reasonable, not all of the path coefficients were statistically significant, indicating potential specification problems within the model.

In response to research question 2a, the standardized path coefficients supported several findings related to the dynamics between programmatic characteristics and outcomes. First, Program Improvement was directly related to Student Impact. Second, Collaboration was positively related to Materials Development, Professional Development, and Program Improvement. Third, Professional Development was positively related to Program Improvement.
In response to research question 2b, three alternative models were specified and estimated given that the overall model fit was not optimal. These alternative models were based on two significant changes: (1) the removal of Project Characteristics and Organizational Practices as exogenous latent variables, and (2) removal of Materials Development. Each of the three alternative models provided a strong fit for the data that was statistically better than the full model fit. When parsimony was considered, it appeared that the most appropriate model for the data was the Collaboration-Impact model without Materials Development (see Figure 12). This model presented a simplified view of the ATE program and indicated that a project's level of collaboration was the critical factor in driving professional development productivity and the level of program improvement.

This study demonstrated that in the context of multisite evaluation, a logic model could be fitted to data collected through an evaluation. Based on this model fit, it was possible to describe the dynamics with which program characteristics contributed to program outcomes. Lastly, it was possible to modify the apriori program logic model and identify a stronger fitting alternative model. These findings demonstrated that SEM can be used as a tool for multisite evaluation. Moreover, by examining alternative models, SEM provides extensive evidence of why program outcomes occur.

The remainder of this Chapter is organized as follows. First, the primary limitations of this study are discussed. Next, the implications of this study for (a) multisite evaluation, (b) theory-based evaluation, and (c) structural equation modeling, are presented. Finally, this chapter concludes with recommendations for evaluators and researchers.
Limitations

There were five primary limitations of this study: (1) sample size, (2) data reliability, (3) measurement models, (4) data types, and (5) a limited program theory. These limitations are presented as the foundation for the implications of this study, which address the fundamental issues that evaluators should consider when using structural equation modeling as a tool for multisite evaluation.

Sample Size

Within the context of this study, the unit of analysis was an ATE project. This unit of analysis was appropriate in the context of the evaluation given that answering the question of why outcomes were produced implied the need to detect project-level effects. In this study the sample size was limited to the number of unique projects that responded to the ATE program evaluation survey and that provided data for each of the latent constructs in the model. These requirements limited the sample to 115 unique projects.

SEM, however, is fundamentally a large sample technique (Kelloway, 1996). Both the estimation methods and the tests of model fit (e.g., RMSEA) are based on the assumption of large samples. A significant body of literature exists on what a minimal sample size should be under different conditions of normality (Boomsma, 1983) and use of different estimation methods (Lei, 2002). Kelloway (1998) indicates that a minimum ratio of 5-10 cases per variable is needed for SEM, which is similar to the frequently cited guidelines for regression analysis. In this study, the sample size was at the low-end of acceptable practice given the number of variables (17) in the study. This was addressed by applying the robust maximum likelihood
estimation method, which simultaneously analyzed the covariance and asymptotic covariance matrices.

Though the impact of the small sample was minimized, this limitation was manifested in the structural model parameter estimates. For example, the standard errors of the path coefficients were large in relation to the path coefficients in the full program structural model (see Figure 8). However, the standard errors were smaller for the more parsimonious alternative models (see Figures 10-12). Similarly, the more parsimonious models fit the data better than did the full program model. This suggests that the primary limitation of a small sample in obtaining a strong model fit is model complexity, i.e., the number of latent and observed variables in the model.

Data Reliability

Collecting consistent data across sites is a critical challenge in multisite evaluation (Hamilton, et. al., 2003). The implied assumption is that data collected prospectively by a central evaluator will be more reliable than data provided retrospectively by different project sites, because the interpretation of the data collection instrument (e.g., a survey) will be consistent across sites. In practice, and certainly in the context of this study, there are multiple steps in the data collection process and the evaluator's involvement varies across those steps. As a result, the reliability of data in a multisite evaluation can suffer.

For this study, the central evaluator designed the data collection instrument with input from ATE projects (The Evaluation Center, 2000). However, the data collection process relied on each individual project to report data based on their interpretation of the survey questions. For example, the survey defined a "module" as "a component that can be used in more than one course" (see Appendix D). This
definition leaves room for interpretation as to what constitutes a "material that can be used in more than one course." As such, it is possible that productivity measures were under or over-reported due to variations in interpretation of the survey questions.

This particular example was manifested by the weak factor loadings for CRSE, MDLE, and OTH_MAT and the breakdown of this component of the measurement model. These results suggest that the interpretation of a "course," a "module," and "other materials" varied across sites.

Overall, this study illustrated that prospective multisite evaluation does not presume reliable data. In prospective scenarios, the individual project still has a critical role in interpreting measures, gathering data, and then reporting to the central evaluator. There is a premium on the evaluator's ability to clearly articulate measures and develop instruments in such a way that projects' interpretations of measures will vary as little as possible.

**Measurement Models**

The definition of the latent variables in the structural model, as manifested in the measurement models was an additional limitation of this study. This study presented two examples that highlighted the limitations of the measurement models in this study. The examples were visible in the following ATE program constructs: (1) project characteristics and (2) materials development.

**Project Characteristics.**

*Project Characteristics* represented the scope and maturity of a project at the time it responded to the survey. This construct was based on the notion that a project that received more money and that was in operation longer would produce
greater results (see Chapter III). The two observable indicators of Project Characteristics were NSF and AGE. NSF was an objective measure that adjusted the total award amount, which was a known quantity, by a factor that represented the proportion of the grant period that had passed at the point in time a project responded to the survey. For example, if a project received a $500,000 award for five years and responded to the survey two years after beginning work, a factor of 40% was applied to the award amount, resulting in a value of $200,000 for that measure. AGE, represented that factor, in terms of years. In this example, the value for AGE would have been 2 years.

The incorporation of the AGE measure into calculating NSF seemed, in retrospect, problematic because AGE is represented twice in the measurement of this construct. The primary indication of this is weak, non-significant factor loadings for both measures, which often reveals collinearity between the measures (Wang, 2003).

An alternative approach might have been to use the total funding amount for the award, rather than a reduced amount that was directly dependent on age of the project. In addition, the project characteristics construct was not comprehensive in describing the variation of contexts within which the projects operated. This construct assumed that all other conditions being equal, more money should produce more results. Of course, we know this was not true—costs vary by geography and technology discipline; equal amounts of money in Ohio and California would not be expected to produce similar results. From the contextual perspective, adjusting total award amounts based on consumer prices or cost of living indices might have provided a better way of measuring the contribution of the program funding.
The second example is the Materials Development construct. For the purposes of this discussion, assume that projects understood the differences between courses, modules, and other materials and that the reporting of this information was accurate. This construct represented materials development productivity by calculating three variable scores based on the number of courses, modules, and other materials under various stages of development. Weightings were applied for the material types as follows: courses were worth 1 point, modules were worth .5 points, and other materials were worth .25 points (see Appendix E). As a result of this weighting scheme, projects were “rewarded” for developing complete courses rather than modules or other materials. Weightings based on the development stage rewarded commercial publication of materials (10 points) over materials in draft stage (1 pt). As a result, projects were also rewarded for dissemination and use.

Several alternatives could be explored to better understand the impact of this measurement scheme. The impact of the alternative could be evaluated by implementing the change, re-estimating the full program model, and comparing the path coefficients of the new and original models. The first alternative is to eliminate one or the other weighting schemes. The second is to aggregate “other materials” and “modules” results into one observed measure for this study. The third is to treat (a) development stages, (b) materials use, and (c) publication as separate metrics.

Notwithstanding the above discussion, it is interesting that the findings related to Materials Development were substantively supported within the overall program context. Recall that materials development productivity was deemphasized by ATE.
over time and that the emphasis of materials development was national dissemination (Hanssen, et. al., 2003). At the same time, the evaluation team reported that actual materials development activity was likely imbedded in professional development and program improvement activity and was therefore not accurately reported on the annual survey (Hanssen, et. al., 2003). This makes sense logically—for example, if a project developed a professional development seminar, it would need instructor and participant materials to support the seminar. These results could be imbedded in professional development activities and under materials development or they could simply be imbedded in professional development, ignoring the actual materials development productivity. As a result, the findings that Materials Development made minimal contributions to Professional Development and Program Improvement could be substantiated given what is known about the operation of ATE and the activities of its various projects.

Data Types

Sinacore and Turpin (1991), Herrell and Straw (2002), and Tushnet (1995) each cited uncontrolled program implementation as a significant challenge for multisite evaluation. In light of this challenge, a significant benefit of uncontrolled multisite program implementation is the opportunity to learn what works under a variety of different program strategies and contexts (Worthen, et. al., 1997). Cluster evaluation, as a variant of multisite evaluation, has as its central focus the exploration of alternative program designs (Herrell & Straw, 2002; W.K. Kellogg, 2000). This study addressed two elements of program implementation by examining the extent of evaluator use and project participation in NSF monitoring activities.
These measures reflected "rigor" in operations and were the basis for the Organizational Practices latent variable.

Another important dimension of multisite evaluation is gathering information about the context within which the project operates (Greenberg, et. al., 2003). Tushnet (1995) asserts that program context is a critical component for identifying why program outcomes occur. This study minimally addressed contextual variables to the degree that project age and funding levels could be regarded as the context for the project. These measures were used to define the Project Characteristics latent variable.

Thus, an important limitation of this study was the lack of information about variations in project implementation practices and project contexts that might have provided additional insights into why program outcomes occurred. These limitations were a product of the program logic model used in this study, which was comprised primarily of intermediate and final program outcomes, and only minimal amounts information about program context and implementation. However, past work has argued that contextual factors (Tushnet, 1995) and implementation strategies (Wang, 1995) also make significant contributions to program outcomes.

While the study successfully fit the structural model to the empirical data, the answers to the question of why outcomes were produced, were limited to the relationships between different levels of outcomes. By not incorporating contextual or implementation measures, formative feedback can not account for contextual variation or for variations in implementation strategy. Given this, a significant purpose of multisite evaluation, to understand how a program operates under varied conditions (Worthen, et. al., 1997) can not be fully addressed.
Limitations of the ATE Program Theory

The overarching theory-related limitation of this study was the limited nature of the program theory that guided this study. In the specific context of this study, the ATE program logic model was articulated by the evaluators (Hanssen, et. al., 2003) after the program had been in operation for 10 years and the evaluation had been underway for 4+ years. The hypothesized relationships between the program components were based on the evaluators' observations of the program and the data that they collected. The impact of this scenario was that the elements of the program model were restricted to the primary survey components. Similarly, the observable indicators of those components were restricted to existing instrumentation.

As a result, the model that was fitted to the data in this study was limited and did not necessarily represent all of the potential operational dynamics within the ATE program. The operational dynamics that were explored addressed (a) how operational rigor (i.e., organizational practices) was related to collaboration, (b) how collaboration contributed to various intermediate program outcomes, and (c) how program improvement contributed to student impact. The contextual dynamics that were explored addressed the relationship between project size and scope (i.e., project characteristics) and collaboration. The strength of these relationships was reflected in the standardized path coefficients that were estimated for the four program models evaluated in this study.

Other potential operational dynamics that could influence program outcomes might include, for example, number of staff, staff qualifications, staff stability, the stability of the project principal investigator, existence of a strategic plan, etc. However, these pieces of data did not exist within the ATE program database, they
were not incorporated into the program model, and they were not part of the structural models that were estimated. Thus, the model parameters could not establish the contribution of these elements to the overall project results.

Ideally, the program logic model used in this study would have been articulated at the outset of the ATE program or, at least before the start of the evaluation. This would have facilitated evaluation activities being keyed to that articulation of program logic.

**Summary of Limitations**

The limitations of this study were due the following five factors: (1) sample size, (2) data reliability, (3) measurement models, (4) data types, and (5) a limited program theory. These limitations underscore the importance of understanding that the application of this method will likely result in a limited response to the question of why outcomes occur within a program. In the specific case of this study, the findings generated through the SEM process represented one possible answer based on one articulation of the ATE program model. Alternative models, some of which were presented in this study, might provide additional answers.

However, the focus of this study was not to uncover a definitive answer to why program results occurred within the ATE program. Rather, the purpose of this study was to evaluate the proposed method for formative evaluation that integrated multisite evaluation, theory-based evaluation, and SEM. The implications of this study for each of these areas are presented below in light of these limitations.
Implications

The purpose of this study was to investigate the viability of an approach for answering the question of why program results occurred that integrated multisite evaluation, theory-based evaluation, and structural equation modeling. This approach was intended to serve a formative evaluation function, which is to help improve programs by providing explanatory evidence about why program outcomes are produced. The multisite context for this study was the ATE program, but generally, multisite evaluation is significant due to the size and scope of many federal multisite programs. Thus, there is a professional imperative to provide the strongest quality evidence in support of program improvement.

The implications of this study are organized according to the three primary concepts for this study: (1) the context of multisite evaluation, (2) the use of program theory to guide evaluation, and (3) the application of structural equation modeling as the analytical engine.

Multisite Evaluation

This study was framed within the general context of multisite evaluation and the specific context of the ATE program evaluation. Previous discussion has shown that ATE was a program that was implemented in an uncontrolled manner and that the evaluation of ATE was conducted prospectively (Herrell & Straw, 2002). In terms of satisfying the definition of multisite evaluation as presented in Chapter II, (a) the object of the evaluation (ATE) was a tangible response to an identified problem, (b) the response was evident in multiple projects at distinct sites, (c) the evaluation of ATE was conducted by a central evaluator, (d) contextual and outcomes data were
collected, and (e) data were pooled across sites to enable comparisons as well as an overall assessment of impact.

The first important implication of this study for multisite evaluation is that the terms prospective and retrospective, as used by Herrell and Straw (2002), are misleading. I would propose a new definition of “prospective multisite evaluation” where (a) the conduct of the evaluation is tightly linked to program implementation, (b) that the evaluator actively collects evaluation data, and (c) that the program theory was developed ahead of the evaluation. A new definition of “retrospective multisite evaluation” would represent the converse of those statements, placing prospective and retrospective multisite evaluations on the opposite ends of a continuum. The third point above that relates to the timing of the program theory development presumes that a program theory will be used in the evaluation. Of course, multisite evaluation does not require the use of program theory, although I would argue that it is a good idea for any evaluator to articulate, at minimum, how they think a program is supposed to work, or is working, before they begin an evaluation of that program.

The second implication of this study is that multisite evaluation should seek three distinct types of data to support three different answers to the question of why outcomes occur: (1) contextual, (2) implementation, and (3) outcomes data. This assertion is an extension of the work summarized in Chapter II, which shows that context and results are the central data elements for multisite evaluation. I would argue that the important implementation variables pertain to the practices employed by projects in pursuit of their goals. For example, an organizational decision to employ trained instructional designers versus relying on existing faculty to develop
courses might be hypothesized as a critical factor in materials development productivity. Similarly, available funding to support a professional instructional designer and/or the institutional context to facilitate hiring such an individual might be hypothesized as important contextual variables.

It should be noted that Tushnet (1995) comes close to articulating the importance of program implementation in multisite evaluation, but her discussion is framed in terms of variation across sites in terms of "activities" in which sites engage. She indicates that "the methods for documenting and evaluating implementation include content analysis of documents, noting, for example, the number of workshops offered and the number of individuals attending" (p. 20). In the context of this study, these activities would have been treated as outcomes.

Thus, the implication of this study is that there are three questions that contribute to the overarching question of why outcomes occur, which might be phrased as follows:

1. How does project context contribute to project outcomes?
2. How do organizational decisions and processes employed by a project contribute to project outcomes?
3. How does productivity on intermediate outcomes contribute to the ultimate project goals?

Each of these questions addresses a component of why program outcomes occur. Within the context of multisite evaluation, data should be collected along each of these dimensions and modeling of a program theory should incorporate these dimensions.
Theory-Based Evaluation

The terms program theory evaluation (Bickman, 1990), theory-based evaluation (Weiss, 1997), and program logic (Funnell, 1997) have been used to describe the practice of basing a program evaluation on a causal model. Chen (1990) argued that program theories can be either normative or descriptive. A normative theory is established apriori and represents how the program is intended to operate. A descriptive theory describes how a program is actually operating and, is often developed after a program has been in operation for some period of time.

This study showed that theory-based evaluation is an ideal compliment for multisite evaluation given that a primary purpose of multisite evaluation is to understand how a program works under a variety of conditions (Worthen, et. al, 1997). Given this, multisite programs represent an opportunity to observe variation in the event, or cause, as well as in the response, or effect (Lipsey, 1993). This study demonstrated that it was possible to enhance the evaluator's understanding of relationships between program components by applying a program logic model to the analysis of evaluation data. At the same time, the implementation of this study implied an ideal model for integrating program theory into a multisite evaluation.

This ideal model can be described as follows. First, the purposes of the evaluation should be both formative and summative. Second, the program theory should be articulated prior to or at the start of the evaluation. Third, the definition of measures and collection of data should be keyed to the program theory. Fourth, evaluation measures should incorporate context, implementation, intermediate outcomes, and final outcomes.
In practice, however, it is apparent a program theory might be articulated at one of three points in time: (1) before or in conjunction with the start of a program, (2) after a program has been operating but before or in conjunction with the start of an evaluation, or (3) after an evaluation has started.

These alternatives also speak to the overall issue of the timing of an evaluation within the context of program implementation. Overall, an evaluation may begin (a) in conjunction with program design and implementation, (b) after the program has been operating, or (c) after the program has ended.

Further, the timing of an evaluation has implications for the general purpose of the evaluation as either formative, summative, or both. Finally, the formative or summative purposes have implications for the definition of measures and collection of data.

In the context of ATE (a) the program theory was developed after the evaluation had started, (b) the evaluation began after the program had been operating, and (c) the purpose of the evaluation was primarily summative. While these conditions are different than the proposed model described above, the conduct of the ATE evaluation serves to highlight these issues.

First, there was a clear attempt to articulate a program model at the beginning of the evaluation. As described in Chapter II, the evaluation team compiled a detailed list of evaluation indicators (see Appendix C) in the early stages of the evaluation (The Evaluation Center, 1999). These indicators were articulated as three levels of drivers, and included process and outcome measures. In contrast to some multisite evaluations (Hamilton, et. al., 2003; Tushnet, 1995), these indicators did not include detailed contextual issues. Also, missing from these initial
stages of the evaluation was a discussion of how these indicators were thought to interact, i.e., articulation of a detailed program logic model. Despite the existence of these indicators, the evaluation data collected ultimately addressed only a portion of these indicators, which included outcomes measures and a limited number of process measures. This emphasis, however, was appropriate given the evaluation questions, three out of four of which were summative (see Chapter II).

This discussion illustrates how (1) evaluation purpose, (2) program theory construction, and (3) evaluation activities interact within a specific evaluation context. I would argue, based on the experience of conducting this study, that these three elements should be aligned. In addition, it is not critical that the evaluation purposes and activities encompass the totality of a logic model. A program logic model can provide a comprehensive framework of how the program is intended to operate by incorporating contextual, implementation, and outcomes measures. The evaluation purposes could be keyed the entire model or to portions of the model, e.g., the levels of outcomes generated. Evaluation activities, e.g., surveys and site visits, though, must be keyed to answering questions articulated by the evaluation purposes and the relevant elements of the program logic model.

The converse, however, would not be recommended. In this case, evaluation purposes and activities would be articulated and then a program logic model would be constructed to reflect only these activities.

The specific context of ATE and this study straddled these two alternatives. First, there was a clear intent to establish a comprehensive set of indicators to guide the evaluation, but a comprehensive logic model was not articulated. Next, evaluation activities were keyed primarily to the summative purposes of the
evaluation and secondly to the formative purpose, which drove the evaluation activity and data collected. Third, a descriptive logic model was developed, which reflected the evaluation activities and data, and only minimally reflected contextual and process indicators. Lastly, the model was tested by fitting the empirical data available, enabling a partial answer to the question of why outcomes occurred.

Restricting the scope of the evaluation to one or two elements of a program logic model has the effect of limiting the evaluator's ability to comprehensively explain why outcomes have occurred. Within the context of multisite evaluation, this may result in the failure to address a key purpose, which is to understand how the program operates under a variety of conditions and implementation strategies.

Structural Equation Modeling

Based on the results presented in Chapter IV, SEM is a promising analysis option to help evaluators understand the relationships between program constructs. This view is supported by authors (Reynolds, 1998; Hennessy, 1999; Wang, 2002) who have used this tool within an evaluation context (though none of the three has taken a detailed look at how SEM could be applied in multisite evaluation context). The future promise of SEM in this context is to expand the range of evaluation questions that are modeled in a multisite context by incorporating examinations of contextual and process measures into a program logic model. This expansion can help evaluators provide a more comprehensive response to the question of why outcomes occur, thus leading to improved formative evaluation.

Overall, the application of SEM in this context depends on (a) appropriate model specification given the available sample size and (b) selecting the appropriate unit of analysis for study.
As stated previously, SEM is fundamentally a large sample technique (Kelloway, 1996). Both the estimation methods and the tests of model fit (e.g., RMSEA) are based on the assumption of large samples. However, sample size requirements present a significant challenge for evaluators in the context of multisite evaluation. As shown in Chapter II, of the five multisite examples reviewed, only ATE presented a sufficient number of sites to support SEM. Even then, the number of projects in the sample was at the low end of acceptable practice given the number of variables in the full program model.

A small sample has important implications for model specification. As demonstrated in Chapter IV, the more parsimonious models specified as alternatives to the full ATE model provided a statistically superior fit to the data. Thus, there was a larger ratio of cases to study variables.

These models were also strongly aligned with the summative purposes of the evaluation. The collaboration models, i.e., those that excluded project characteristics and organizational practices, modeled the relationships between program outcomes only, i.e., these models eliminated contextual and process measures. This suggests that an approach to using SEM in multisite evaluation is to constrain the number of parameters to be estimated to a number that can be supported given the available sample size. In addition, those parameters should represent relationships within the model that are keyed to specific evaluation purposes and that represent a portion of the overall program theory. This practice will have the effect of limiting responses to the question of why outcomes occur to the estimable parameters. A benefit to this approach, however, may be that parameter estimates are more stable. This is exactly what occurred in this study, when the alternative models were evaluated.
An alternative approach is to expand the sample size by collecting data from program *participants* at individual sites. The opportunity to increase sample size is cited as one of the key benefits of conducting multisite evaluation (Sinacore & Turpin, 1991). This approach however, may only be appropriate when the program goals are defined in terms of participant-level metrics. This is the scenario described by Greenberg, et. al. (2003), who examined welfare to work programs where the primary outcome metric was defined as change in participant income, and Reynolds (1998) where the primary outcome measures were student achievement scores. In the specific context of ATE, adopting this approach would have required modifying outcome measures to focus on student-level goals. Hypothetically, these measures might have included (a) time to complete a program, (b) time between completion and full-time employment, and (c) supervisor ratings of on-the-job performance. The implication for evaluators, however, would have been a much different approach to constructing measures and collecting data. Rather than focusing data collection at the project level, data collection would have focused at the student level.

This issue, however, raises a fundamental concern about the requirements for establishing causation. As Lipsey (1993) stated, it is important that both the event and the response vary. If the event/treatment is constant, but the responses vary, then it is impossible to determine the cause of the response. In the context of a multisite program evaluation, participant data represent one sample size, whereas the site-level data, which describe the treatments (context and process), will have a much smaller sample size. This situation describes a nested research design with students nested within projects. Thus, if the goal of the evaluation is to answer the
question of why outcomes occur in terms of the program, the sample size is necessarily limited to site-level measures.

The implication for applying SEM in the context of multisite evaluation may be that the opportunities will be limited to large programs, where data can be acquired from a sufficient number of sites in relation to the number of parameters being estimated. In the federal multisite program context, this is probably achievable, but this approach will not be appropriate for all programs.

Summary

The following summary points address what I view as the key lessons for integrating multisite evaluation, theory-based evaluation, and structural equation modeling with the intent of providing formative evaluation feedback.

1. *Multisite program theory should incorporate context, process, and results, to comprehensively address why outcomes occur.* Within the context of multisite evaluation, this expands the premise that contextual and outcomes data are important. The critical element of this assertion is to emphasize operational practices as the key process elements, rather than choices between competing sets of activities.

2. *Development of evaluation measures and instruments should be directly keyed to a stated program model.* Where possible, stated program theory should drive the development of measures, rather than theory being derived based on available data.

3. *Theory-based evaluation and structural equation modeling are complimentary approaches that can be applied to multisite evaluation contexts.* The basis of this argument is that both theory-based evaluation and SEM presume an
interest in determining the causes and effects within a program structure with an eye towards program improvement. A central purpose of multisite evaluation is the desire to learn what works in a variety of contexts. The alignment of these purposes supports the combined use of these methods.

4. **Structural equation modeling is appropriate in cases where a sufficient number of sites exist to support model estimation.** This presumes that the site, not individual participants will be the unit of analysis for the evaluation.

5. **Structural equation modeling can be used to address program contexts, processes, and outcomes.** Evaluating a portion of the program logic model limits the response to why outcomes occur. Many evaluations simply cannot address all of the potential elements that contribute to program outcomes. Choices are made and the components of program theory that are emphasized are keyed to the objectives of the evaluation. Thus, any response to the question of why outcomes occur will be one response of many possible alternatives.

**Recommendations**

The following are recommendations for evaluators, who might consider this method for the next multisite evaluation they conduct, and for researchers who are interested in adding to the body of knowledge on multisite evaluation, theory-based evaluation, and structural equation modeling.

**Recommendations for Evaluators**

1. **Begin the evaluation of multisite programs with a clear articulation of the program theory and use that theory to clearly develop measures for each program component.** This recommendation argues for a normative program model, but given
that many evaluations begin after the program has been operating for a period, constructing a descriptive model before embarking on data collection is recommended.

2. **Design evaluation activities around a prospective approach to collecting data.** Where it is impractical to directly collect data from a site (e.g., due to the number of sites), develop measures that are easily interpretable and for which data are easily collected and reported. While many multisite evaluations will rely to some degree on the self-report data from projects, the degree to which this issue can be mitigated will enhance the credibility of results and evaluative feedback.

3. **Collect site level data to support SEM.** Collect participant level data in cases where program goals can be clearly articulated in terms of participant impacts. Multisite evaluation is concerned with comparability across sites and the aggregation of site results. Applying SEM in this context should be keyed to site-level metrics. In cases where program outcomes can be articulated in terms of participant results, and data can be gathered from participants, then these data should be aggregated within a site to generate a site-level measure of participant impact.

4. **Ensure that the unit of analysis for SEM is appropriate given the evaluation design and the primary program outcome measures.** As an example, if the program outcomes are reported in terms of a site, then do not use individual student-level data as the basis for analysis. If student-level data are used, they should be aggregated at the site level and reported as a site level measure.

5. **Standardize the latent variable indicators to mitigate issues associated with variable scales.** Normalize the data when multivariate normality is not present. Given the small samples sizes that are likely with SEM in a multisite context,
standardization of variables will facilitate the model estimation and fitting process. While this technique will also make interpretation of path coefficients more difficult, this should be less of a concern as the primary interest is in the relationships between latent constructs. For example, if we know that high levels of A predict high levels of B, and that B is a desired result, then direct feedback for improving A can be formulated. Similarly, normalizing data will help satisfy the assumption of multivariate normality for SEM.

6. Validate SEM results through cross validation with a comparable sample.

In a multisite context, it may be difficult to acquire a sufficient sample size to cross validate model estimation and overall fit results. Developing a longitudinal design, assuming that point-in-time is an important program factor, may mitigate this constraint.

Recommendations for Researchers

1. Test the assertion that site-level data are preferred within the context of multisite evaluation, limiting the opportunity to increase sample size by using participant data. This assertion could be examined via a simulation where observed exogenous variables are restricted to a limited number of categories and observed endogenous variables are allowed to vary freely. The simulation would test the hypothesis that parameter estimates and overall fit statistics for the site-level model are more robust than for the model based on participant level data. A primary consideration in this study would be the evaluation of fit statistics that are not sensitive to sample size.

2. Model the ATE program under different definitions of materials development. Based on the results of this study and the subsequent discussion of
findings, there are three possible reasons for the minimal impact of materials development within the model: (1) poor measurement of the construct, (2) inconsistent interpretation of the metrics by the projects, and (3) substantive program changes reflecting the de-emphasis of materials development. This study would reduce the possibility that the first reason is the primary contributing factor for the findings in this study.

3. Model the ATE program by incorporating additional contextual measures that better represent the size and scope of individual projects. The addition of more contextual measures would enhance the understanding of contributions of different project contexts to project outcomes. Conversely, such a study may validate the findings from this research that indicated there was no relationship between project characteristics and project results.

4. Model the ATE program at different points in time. Research in this area would contribute to an improved understanding of project context by determining an effect due to project longevity and/or calendar year of operation, i.e., is there a difference between projects that operated pre-2000 to those that operated after 2000? The challenge in this process would be achieve sufficient sample size within groups to obtain reliable parameter estimates.
REFERENCES


APPENDIX A

HSIRB Approval Letter
Date: February 12, 2004

To: Brooks Applegate, Principal Investigator
    Arlen Gullickson, Co-Principal Investigator
    Carl Hanssen, Student Investigator for dissertation

From: Mary Lagerwey, Ph.D., Chair

Re: HSIRB Project Number 04-02-01

This letter will serve as confirmation that your research project entitled “Structural Equation Modeling as a Tool for Multisite Evaluation” has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: February 12, 2005
August 25, 2003

Mr. Carl Hansen
4701 Sunflower Ridge Drive, N.E.
Ada, MI 49301

Dear Carl:

This letter is in response to your August 8 request for permission to use the ATE evaluation 2000-2003 survey database as a basis for your doctoral dissertation. As I stated on August 8, you have permission to use these survey data. This permission carries several provisos:

1. You need to comply with the University's human subjects requirements for dissertations and for use of data.

2. You note in your letter, and I confirm here, that you must protect the confidentiality of these data. Protection of this confidentiality includes the following pieces:
   - No data are to leave the project database system in The Evaluation Center without first being stripped of all codes that identify individual projects.
   - Any person, such as Dr. Brooks Applegate, who will have access to the data set should provide a written statement that confidentiality of these data will be protected and that the data shared with such person will not be copied and used outside the confines of this study that you are conducting.

3. Findings from your study will be made available to the ATE project and can be published as part of the ATE results on the ATE Web site or other venues as determined by our ATE project staff. Such use by our project will be done in concert with you and after discussion with you on this matter.

I am pleased that you have chosen to use the data set for your dissertation work. We will do our best to assist you in conducting this effort. We look forward to your findings and using them to serve ATE needs.

Sincerely,

Arlen R. Gullickson
Project Director
APPENDIX C

ATE Evaluation Indicators

127
Level I: Initiation and Support of Projects and Centers

NSF funds those projects and centers that will most effectively help reach ATE program goals and monitors, supports, and evaluates projects and centers.

Driver 1: Strategic Planning

Enabling Outcome: An effective strategic plan exists, is being implemented, and is conveyed clearly to others.

Indicators:

1. Is the strategic plan linked to goals identified by enabling legislation?
2. Does the plan include a conceptual model or rationale for how the desired outcome would be achieved?
3. To what degree do RFPs and grant applications specify program priorities?
4. To what degree do ATE communications inform potential and new projects and centers of their responsibilities for long-term viability (i.e., sustainability, dissemination, institutionalization)?
5. Is a strategic plan for ATE available for review?
6. How "user-friendly" is the plan (i.e., are the goals, objectives, priorities, etc. clear)?
7. To what degree do appropriate ATE publications and web sites outline priorities?
8. Is the strategic plan dynamic in response to information from the field?

Short-Term Outcome: Most grant applicants are aligned with the ATE program's objectives and priorities and all funded projects and centers are aligned.

Indicators:

9. To what degree are the goals and objectives of projects and centers aligned with the goals and objectives of the ATE program?

Driver 2: Funding

Enabling Outcome: An effective review protocol exists, is utilized, and is based on program strategy/priorities.

Indicators:

1. Does an effective proposal review protocol exist?
2. To what degree is it consistent with the strategic plan/program priorities? (Do reviewers have an
appropriate background? Are the criteria for the review process consistent with program guidelines?)

3. To what degree is the proposal review protocol followed?

4. Are reviewers adequately prepared? Are participants of the proposal review process clear about criteria and process for reviews?

5. Do proposal review teams meet in full and on time?

6. Are individual reviews written for each proposal?

7. Is a review summary written for each?

8. Are preaward site visits made for centers and major projects?

9. Is negotiation of project and center focus and award amount effective and timely?

10. Does NSF bring people together from programs with similar goals? (Avoid reinventing the wheel. Promote synergism for both content and process.)

**Short-Term Outcome:** Review team and program officers make funding decisions based on clear priorities and protocol.

**Indicators:**

11. What percentage of project and center goals and objectives align with program goals and objectives?

12. What percentage of awards and award dollars are made to the top ATE program priorities in terms of
   1. technology field
   2. product type
   3. student population targeted
   4. viability of work

**Driver 3: Monitoring**

**Enabling Outcome:** Resources are available, procedures are in place, and requirements are clear for monitoring project and center activities.

**Indicators:**

1. To what degree does the ATE program have clear-cut requirements for monitoring centers and projects?

2. Are sufficient staffing and monetary resources provided to effectively monitor centers and projects and make recommendations?

3. To what degree does ATE management follow up to ensure that project monitoring occurs as planned?
4. Do ATE staff review all annual and final reports and other submissions?
5. Does ATE follow up with centers and projects based on submitted reports?

**Short-Term Outcome:** Projects and centers are adequately monitored.

**Indicators:**

6. To what degree does NSF follow its plan and monitor project and center progress?
7. To what degree does NSF respond to project and center annual reports?
8. Under what conditions are site visits made to projects and centers?
9. What are the intentions and results of site visits?

**Driver 4: Supporting**

**Enabling Outcome:** Resources are available, procedures are in place, and requirements are clear for supporting project and center activities.

**Indicators:**

1. To what degree does the ATE program have clear-cut requirements for supporting centers and projects?
2. Are sufficient staffing and monetary resources provided to effectively support centers and projects?
3. To what degree does ATE management follow up to ensure that support activities occur as planned and make new centers aware of these activities?
4. Are resources available to offer nonmonetary support regarding identified needs to projects and centers?
5. Is the monitoring process understood by center and project personnel?
6. Are mechanisms in place to react to support level and activity level changes?

**Short-Term Outcome:** Projects and centers are adequately supported.

**Indicators:**

7. To what degree does NSF support projects and centers when specific needs are identified (e.g., help with dissemination or PI training)?
8. To what degree does the ATE program assist projects and centers with long-term viability (dissemination, sustainability, transportability)?
Driver 5: Evaluating

Enabling Outcome: Resources are available, procedures are in place, and requirements are clear for evaluating project and center activities.

Indicators:

1. Are evaluation requirements clear in grant proposal forms and other documents?
2. Is it clear how NSF plans to use project and center evaluation reports?
3. Does NSF follow up with those projects and centers that do not submit a plan for evaluation?

Short-Term Outcome: Project and center activities are adequately evaluated.

Indicators

4. Do projects and centers complete evaluations per NSF specifications?
5. How does NSF communicate evaluation requirements?
6. Does NSF receive evaluation reports from projects and centers?
7. In what way does NSF use the reports?
8. To what degree is the need for professional evaluation made clear to projects and centers?
9. To what degree does NSF monitor evaluation?
10. Does NSF formally evaluate the ATE program every 3 years (3 years is the average length of an award)?

Level II: Development

Short-term outcomes' impact at the student and faculty levels.

Driver 1: Collaboration

Enabling Outcome: A thoughtful plan and process exist for identifying and bringing potential collaborators together.

Indicators:

1. Is there a clear purpose for the collaboration?
2. Is there a system for identifying all potential collaborators?
3. Is there a system for determining a viable basis for collaboration with potential collaborators?
4. Are the collaborations characterized by formal agreements?
5. Are reasonable meeting schedules/methods available?
6. Is there an organizational/management structure in place that would support collaboration?
7. Were potential collaborators supported through initial stages of collaboration?
8. Are purpose and program clear?
9. Are role, contribution, division of labor, and comparative advantage clear?
10. Are all roles and capacities covered?
11. Are all work products and deliverables set on a time frame?
12. Are levels of investment, risks, and returns on investment understood?
13. Is there CEO approval and organizational support?
14. Is there equality of voice (ethics/conflict of interests)?
15. Is communication protocol agreed upon?

Short-Term Outcome: Collaborations are effective and have become "institutionalized" or self-supporting.

Indicators:

16. With whom do you collaborate?
17. What purposes do your collaborations serve?
18. Are meetings supported by the members of the collaborating groups?
19. Are collaborating groups developing their own goals and agendas?
20. Are new members joining collaborating groups?
21. Are new collaborations being formed?

Driver 2: Standards Development

Enabling Outcome: Are there procedures in place that would facilitate use or development of work force standards or guidelines?

Indicators:

1. Have current and future work force needs been investigated?
2. Has an effective process been put in place for creating or refining industry-based standards or guidelines?
3. Has a coherent set of knowledge and skills been developed?
4. Are there adequate plans for verifying the standards?

Short-Term Outcome: Have the standards been usefully applied in a variety of settings?
Indicators:

5. How have the standards been used by industry?
6. Have the standards been used for curriculum development?
7. Have the standards been used to guide program development?
8. Have the standards or guidelines been used to inform faculty development?

Driver 3: Curriculum

Enabling Outcome: Are there processes in place that would ensure the development of quality curriculum, instruction, and assessment (CIA)?

Indicators:

1. Are experts involved (curriculum development specialists, content specialists, SMET education specialists, assessment specialists, industrial specialists, SCANS specialists)?
2. Are adequate facilities available (instructional technology, printing, laboratories, etc.)?
3. Is there access to world class and pertinent industrial technology?
4. Is there a system for pilot testing and field-testing developments and incorporating feedback?
5. Is there a set of universal standards or minimums for CIA? (How is the quality of development measured?)
6. Are individuals at transfer institutions involved in CIA development?

Short-Term Outcome: Are the curriculum, instruction, and assessment aligned and reflective of cutting edge or world class work force need?

Indicators:

Industry-Related:

1. Is the CIA industry verified? (Has there been a thorough evaluation of current industry needs? Have the cutting edge needs been identified? Was industry involved in identifying those needs?)
2. Is there a system in place for continuing assessment of identified needs?
3. Is the CIA and context for learning industry-based?
4. Are industry representatives involved in assessment?

**Basic Skills-Related:**

5. Does the CIA provide a strong foundation for SMET?
6. Is the CIA logically organized and aligned?
7. Does the CIA address nontechnical skills, e.g., communicating effectively?

**Learning-Styles-Related:**

8. Does the CIA match the needs of the people taking the course? Is it inclusive? Does it allow for different learning styles?
9. Is the CIA inquiry/problem-solving based? Does it promote student-centered learning (e.g., community of scholars, hands-on, minds-on)?
10. Is the CIA cutting edge?
11. Is the CIA group-based?

**Articulation-Related:**

12. Is the CIA in line with educational standards required for continued or advanced education?

**Assessment-Related:**

13. Are students given formal feedback on their performance?
14. Are students involved in their own assessment?
15. Is assessment information used to improve the CIA?
16. Do students exhibit the skills that were intended by the curriculum and instruction?

**Driver 4: Program Development**

**Enabling Outcome:** Are there processes in place that would ensure quality program development?

**Indicators:**
1. Were experts involved in program planning?
2. Are there adequate facilities and resources for the program?
3. Were power players involved?
4. Was an appropriate program architecture identified?
5. Was CIA analyzed for resource needs?
6. Did institutions provide for program sequence following standards?
7. Do institutions support "fast track" changes in CIA and programs?
8. Does the program meet accreditation standards?

**Short-Term Outcome:** Does the program reflect cutting edge or world class work force needs?

**Indicators:**

9. Are students graduating with world class skills? Are students being employed in world class or locally comparable positions? To what extent? At what level of entry?
10. Has the project or center won awards? What recognition exists for the program (e.g., certification, accreditation, recognized award, accepted by field, etc.)?
11. To what extent are programs/courses articulated or aligned for 2+2+2?
12. Are the program systems efficient?
13. Does the degree program match the courses? Does the degree program reflect wise placement of courses?

**Driver 5: Professional Development**

**Enabling Outcome:** Are there effective, ongoing opportunities for professional development?

**Indicators:**

1. Is the content aligned with the needs of industry? Is there an industry advisory committee?
2. Do faculty interact with or work in industrial settings that parallel the program?
3. Are summer or "out of term" training sessions provided?
4. Is mentoring provided?
5. Are opportunities available for reflection on and discussion of practice?
6. Are external experts or expert materials used or available?
7. Is there a method for determining faculty strengths and needs prior to participation?
8. Have the faculty won awards?
9. What skills have faculty gained through participation?

Short-Term Outcome: Are "developed" faculty of world class quality and committed to continuous improvement?

Indicators:

10. Do faculty request professional development opportunities that are ongoing over the course of their career?
11. Do faculty and industry support and/or interact with each other?
12. Do students and experts (educational, SMET, and industrial) view the faculty as world class?

Driver 6: Recruitment Activities

Enabling Outcome: Are complementary processes in place to reach out to diverse types of students?

Indicators:

1. Is there an effective recruitment plan?
2. Does the plan include how to reach young students (high school or junior high)?
3. Does the plan include how to articulate from a secondary to a 2-year institution?
4. Does the plan include how to reach persons already in the workforce including skills upgrade?
5. Does the plan include how to reach different cultural groups, genders, races, and nontraditional students?
6. Does the plan include how to reach different SES groups?
7. Does the plan work with displaced workers and unemployment agencies?
8. Does the plan reach international groups?

Short-Term Outcome: Are the types of students applying to the program reflective of the U.S. population?

Indicators:
9. Do numbers of students applying from various cultural groups and genders match the U.S. population?
10. Are the numbers of applicants moving toward matching the number of the groups in the U.S. population?

**Driver 7: Student Services**

**Enabling Outcome:** Are students provided the services they need to help them be successful?

**Indicators:**

1. Are students provided opportunities for developmental studies in areas of need?
2. Are students helped to develop socialization and study skills?
3. Are employability skills incorporated into the program (including leadership skills)?
4. Are there opportunities to socialize with people in the same industry?
5. Does Student Services intend to fully coordinate the efforts of counselors at all levels?
6. Are students provided interpersonal support?
7. Are counselors aware of the importance of the project, and do they support its goals?
8. Are students provided tutoring assistance?
9. Are students provided appropriate internship experiences?
10. Are students provided job placement assistance?

**Short-Term Outcome:** The types of students graduating from the program reflect the U.S. population.

**Indicators:**

11. Do numbers of students graduating from various cultural groups and genders match the U.S. population?
12. Are the numbers of graduates moving toward matching the number of the groups in the U.S. population?

**Level III: Viability of Project and Center Work**

*Long-term outcomes impact at the workforce level.*

**Driver 1: Sustainability**

**Enabling Outcome A:** A process is in place to encourage and support institutionalization.

**Indicators:**
1. Administrative support has been sought to begin the process.
2. Administration is kept informed of project or center progress.
3. Regular meetings are arranged between project or center staff, colleagues, and administrators.

**Short-Term Outcome A:** The project or center work and/or products become institutionalized.

**Indicators:**

4. The new course, program, or technique becomes part of regular course offerings.
5. Staff are given enough release time to continue the work.
6. Staff work becomes part of their job descriptions.
7. Adequate resources exist to support continued work.

**Enabling Outcome B:** A process exists and is being implemented to secure continued support for the long term.

**Indicators:**

8. A funding plan exists.
9. Staff is available to carry out the plan.
10. Potential funders have been identified.
11. Potential funders have been contacted.

**Short-Term Outcome B:** The project or center work becomes supported by non-NSF funds.

**Indicators:**

12. Other organizations, associations, etc. begin providing greater amounts of financial resources.
13. The work continues, but no additional NSF grants are sought.

**Long-Term Outcome:** To what extent does the work of the project or center continue and grow?

**Indicators:**

14. Has there been a change in staffing level?
15. Has there been a change in goals/objectives?
16. Do long-term plans exist?
Driver 2: Transportability

Enabling Outcome: A plan exists and is used to ensure that products are transportable across the nation.

Enabling Outcome B: A plan exists and is used to ensure that products are transportable across technology fields where possible.

Indicators for A and B:

1. Have transportability issues been investigated?
2. Have findings been integrated into project or center work?
3. How transportable will this be in the future?

Short-Term Outcome A: Project or center products could be used with little or no modification in other parts of the country.

Short-Term Outcome B: Project or center products could be used with little or no modification in other technology fields.

Indicators for A and B:

4. To what degree can your products be used in different locales across the nation without requiring major modification?
5. To what degree can your products be used by other technology fields without requiring major modification?
6. Is the program available in a foreign language?
7. How easily can the program be modified by off-site users?

Long-Term Outcome: Where possible, project and center products are transferrable to regions and fields.

Indicators:

8. To what degree have products been requested by/used in other regions?
9. How much modification is required?
10. To what degree are products requested/used by other tech fields?
11. How much modification was required?
12. How successful was implementation of the program at the transfer site?

Driver 3: Dissemination

Enabling Outcome: Procedures are in place to encourage and facilitate dissemination of project and center work and products.
Indicators:

1. Is there a plan for dissemination?
2. Are resources and staff available to conduct dissemination activities?

Short-Term Outcome: To what extent and in what ways do the projects and centers disseminate their work and products?

Indicators:

3. To what extent is the dissemination plan being carried out?
4. To what extent do projects and centers work together to disseminate?
5. What are the dissemination activities for projects and centers?
6. What are their costs?
7. Who are the targets of dissemination activities?
8. Who are the actual recipients/participants?
9. To what degree do activities lead to
   1. Dissemination of product information
   2. Dissemination of products without actual use in classroom
   3. Dissemination of products with application in classroom

Long-Term Outcome A: To what extent is the collective dissemination impacting technician education and the work place?

Indicators:

10. How many non-ATE-awarded institutions have accessed ATE products?
11. How many have used them?
12. How many plan to continue using them?
13. Any reported results from use?

Long-Term Outcome B: To what extent is the work force impacted?

Indicators

14. Have the numbers of workers in the advanced technology work force increased?
15. Has the fit between work force needs and worker knowledge and skills improved?
16. Are more people from underrepresented groups part of the advanced technology work force?
APPENDIX D

Survey Instrument
Basic Information—Required

This section provides basic information about your center/project ATE grant. Please verify or correct the provided information and complete where needed.

Numerical values must be entered as integers (e.g., "3420", "6" or "0"). Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").

Basic Information about Your Center/Project
1. a. Funded Institution:

b. Institution Category: Place an X next to Only One.
   __ 4 year college/university
   __ 2 year college
   __ Association/Society
   __ Secondary School
   __ Other

c. Funding Category: Place an X next to Only One.
   __ Project
   __ Center
   __ Articulation Partnerships

d. Begin date of current NSF-ATE funding: __/__/____ (MM/DD/YYYY)

e. End date of current NSF-ATE funding: __/__/____ (MM/DD/YYYY)

f. Current award amount: $__________

g. Project Director / Principal Investigator
   Title: Place an X next to Only One.
   __ Dr. __ Mr. __ Mrs. __ Ms.
   Contact Information: Please complete.
   First Name: _____________________________________________
   Middle: ________________________________________________
   Last Name: ______________________________________________
   Email: __________________________________________________
h. Technology Field: *Place an X next to Only One.*

- A. Agriculture
- B. Aquaculture
- C. Biotechnology
- D. Chemical Technology
- E. Distance Learning
- F. Electronics, Instrumentation, Laser and Fiber Optics
- G. Engineering Technology (general)
- H. Environmental Technology
- I. Geographic Information Systems
- J. Graphics and Multimedia Technology
- K. Information Technology, Telecommunications
- L. Machine Tool Technology, Metrology
- M. Manufacturing and Industrial Technology
- N. Marine Technology
- O. Mathematics
- P. Multidisciplinary or Interdisciplinary (General)
- Q. Physics
- R. Semiconductor Manufacturing
- S. Transportation
- T. Other
PI Overview—Required

This section should indicate the Principal Investigator's (PI's) view of the Center/Project and reflect information provided in the other sections of this survey. Unless indicated otherwise, please fill out every question and items in these questions. Thank you!

Numerical values must be entered as integers (e.g., "3420", "6" or "0"). Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").

The Principal Investigator's Overview of the Center/Project

I. Time and Status

I. Where is the project/center in its life cycle? Place an X next to Only One.
   __ 1-3 years
   __ > 3 years
   __ Other. Please describe ________________________________

II. Is this the last year of the project’s/center’s work? Place an X next to Only One.
   __ Yes
   __ No
   __ Other. Please describe ________________________________
III. Please rate the current status of your center/project as compared to its status last year at this time for each of the following factors. For each item a-i, place an X under Only One of the 6 rating options (e.g., Stable).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Not Applicable</th>
<th>Substantial Decline(&gt;20%)</th>
<th>Some Decline(5-20%)</th>
<th>Stable</th>
<th>Some Increase(5-20%)</th>
<th>Substantial Increase(&gt;20%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Number of collaborations (relationships with institutions or groups that provide money and/or other support)</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>b. Financial support from other organizations</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>c. Use of center/project-developed products</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>d. Participation in project/center activities by other institutions and organizations</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>e. Students enrolled</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>f. Students graduating or completing the program</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>g. Students placed in related technical jobs, whether they completed program or not</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>h. Number of professional development opportunities (e.g., conferences, workshops, inservice, on-line courses)</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
<tr>
<td>i. Number of participants in professional development opportunities</td>
<td>______________</td>
<td>________________________</td>
<td>__________________</td>
<td>______</td>
<td>__________________</td>
<td>______________________</td>
</tr>
</tbody>
</table>

2. If you conducted a workforce needs assessment in the last 12 months, please CHECK ALL METHODS that you used. If you place an X next to Not Applicable, please go to 3. Otherwise, place an X next to all that apply.

- This question is Not Applicable
- Survey
- Review of existing reports or other literature
- Interviews
- Focus groups
- Other. Please describe

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
3. Center/Project Evaluation

a. If you have an evaluator, is/are the evaluator(s) (choose one). *Place an X next to Only One.*
   - This question is Not Applicable
   - External (hired specifically to evaluate this grant)
   - Internal (is a member of center/project staff)
   - Both (you have both types of evaluators)

b. How useful is your project’s/center’s evaluation to your project? (choose one). *Place an X next to Only One.*
   - This question is Not Applicable
   - Not useful
   - Minimally useful
   - Some use
   - Useful
   - Essential to the project/center

c. If applicable, describe in what way(s) you used the evaluation in your project/center.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

d. To what extent do your project’s/center’s evaluation findings provide evidence of the quality of your outcomes? (choose one). *Place an X next to Only One.*

   - This question is Not Applicable
   - No evidence
   - Some evidence
   - About half of the evidence
   - Most of the evidence
   - All of the evidence
4. Collaboration: If you collaborate with other ATE projects/centers, please CHECK ALL THAT APPLY (If you place an X next to Not Applicable, please go to 5. Otherwise, place an X next to all that apply).

We collaborate for:

- This question is Not Applicable
- Materials development
- Professional development (e.g., workshops)
- Best practices development
- Sharing of project/center products
- Sharing of best practices
- Other. Please describe ________________________________

5. Product dissemination: Indicate what method(s) your center/project uses to disseminate your center/project's products regionally or nationally by CHECKING ALL THAT APPLY (If you place an X next to Not Applicable, please go to 6. Otherwise, place an X next to all that apply).

- This question is Not Applicable
- In-house publication and distribution
- Commercial publication
- Presentations at regional/national conferences or meetings
- Web page
- Other (please describe) ________________________________

6. Please CHECK ALL STEPS THAT APPLY for how you recruit and/or retain for the ATE-grant funded program (If you place an X next to Not Applicable, please go to 6II. Otherwise, place an X next to all that apply).

- This question is Not Applicable
- Written materials (e.g., brochures, newsletters)
- Web sites about the program
- Presentations by invited speakers
- College fairs at secondary schools or other locations
- Campus visit programs
- Summer or academic workshops for students (e.g., STEM or technician-skill development, career awareness)
- Summer or academic year workshops for teachers
- Work-related experiences for students (e.g., day on the job, visit to business, internship)
- Targeted workshops
- Financial aid (e.g., scholarships, work study)
- Tutoring
- Articulation agreements
- Counseling
- Other. Please describe ________________________________
II. Please CHECK ALL STEPS THAT APPLY specifically for recruiting and/or retaining underrepresented groups (e.g., minorities, women, people with disabilities) for the ATE-grant funded program (If you place an X next to Not Applicable, please go to 7. Otherwise, place an X next to all that apply).

- This question is Not Applicable
- Written materials (e.g., brochures, newsletters)
- Web sites about the program
- Presentations by invited speakers
- College fairs at secondary schools or other locations
- Campus visit programs
- Summer or academic workshops for students (e.g., STEM or technician-skill development, career awareness)
- Summer or academic year workshops for teachers
- Work-related experiences for students (e.g., day on the job, visit to business, internship)
- Targeted workshops
- Financial aid (e.g., scholarships, work study)
- Tutoring
- Articulation agreements
- Counseling
- Other. Please describe

7. If applicable, please describe your placement strategies employed for the ATE-grant funded program. Optional question.

8. Please CHECK THE TOP TWO (IF APPLICABLE) regarding what you believe are the most important ways in which classrooms and other educational experiences for students have changed as a result of your center’s/project’s work (If you place an X next to Not Applicable, please go to 9. Otherwise, place an X next to the top two that apply).

- This question is Not Applicable
- Increased use of work-based skills in curricula
- Increased interest in learning by students
- More relevant and up-to-date materials available
- Movement away from traditional lecture delivery of lessons
- Other. Please describe

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
9. For any significant unintended outcomes (positive and/or negative) of your center/project work, please CHECK ALL THAT APPLY (If you place an X next to Not Applicable, please go to 10. Otherwise, place an X next to all that apply).
   __ This question is Not Applicable
   __ Partnerships, networks, collaborations (i.e., relationships with institutions or groups that provide money and/or other support) increased beyond those planned
   __ Applications to or work for other disciplines occurred
   __ Additional funding received
   __ Loss of staff to business opportunities
   __ Communication or work-related difficulties with collaborating partners
   __ Other(s). Please describe ____________________________________________

10. Please provide up to three barriers or challenges to success that occurred in your center/project. Optional question.
   a. Barrier #1:
      __________________________________________________________________________
      __________________________________________________________________________
      __________________________________________________________________________

   b. Barrier #2:
      __________________________________________________________________________
      __________________________________________________________________________
      __________________________________________________________________________

   c. Barrier #3:
      __________________________________________________________________________
      __________________________________________________________________________
      __________________________________________________________________________
11. Advisory Committees

a. If you have advisory committee(s) to serve the project’s/center’s needs, CHECK ALL THAT APPLY for committee type (If you place an X next to Not Applicable, please go to 12. Otherwise, place an X next to all that apply).

___ This question is Not Applicable
___ Local institution or other locally based group
___ Regional
___ National (e.g., National Visiting Committee)
___ Other. Please describe _____________________________________________________

b. If applicable, please describe the activities of your advisory committee(s). Optional question.

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________

12. Describe your plans for sustainability, if any, of your project/center. Optional question.

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________

13. Please describe any other important features of your center/project that are not captured in the survey. That is, what center/project features would you like to highlight that have not been described elsewhere. Optional question.

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________

14. What is your view of the effectiveness and value of the ATE program? Optional question.

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
Questions 15 a-d are Optional

15. a. What features of the survey (e.g., web interface) did you find most helpful?
   
   b. What features of the survey should be changed?
   
   c. How much time, including data collection and on-line time, did it take you to complete the survey this year?
   
   d. Additional comments regarding the survey itself.

16. What aspects of your project/center are likely to be institutionalized (i.e., remain in the institution after the project/center has ended)? Optional question.
17. Articulation Agreements

a. If applicable, what type of articulation agreement(s) have been established through your project's/center's work? Please CHECK ALL THAT APPLY.

- This question is Not Applicable
- Between secondary and 2-year colleges
- Between secondary and 4-year colleges
- Between 2-year and 4-year colleges and universities
- Other. Please describe.

b. What are the purposes/focuses of these agreements? Please CHECK ALL THAT APPLY.

- This question is Not Applicable
- Strengthen the technological knowledge and skills of K-12 teachers
- Strengthen the science and mathematics preparation of K-12 teachers
- Facilitate the transition of students from STEM (Science, Technology, Engineering, Mathematics) associate's degree programs to related bachelors degree programs, especially those having a strong technological basis.
- Other. Please describe.
Monitoring—Required

Confidentiality of responses to this section will be provided to the extent allowed by law. *Unless indicated otherwise, please fill out every question and items in these questions. Thank you!*

Numerical values must be entered as integers (e.g., "3420", "6" or "0").
Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").

### NSF Monitoring of Centers and Projects

1. Indicate the frequency of the following monitoring actions between your center/project and your NSF program officer during the past 12 months. *For each item a-f, place an X under Only One of the 4 Frequency options.*

<table>
<thead>
<tr>
<th>Monitoring Action</th>
<th>Frequency (Number of Times)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>a. Site visits</td>
<td></td>
</tr>
<tr>
<td>b. Telephone calls</td>
<td></td>
</tr>
<tr>
<td>c. Email contacts</td>
<td></td>
</tr>
<tr>
<td>d. Visits to NSF</td>
<td></td>
</tr>
<tr>
<td>e. Principal Investigator meetings</td>
<td></td>
</tr>
<tr>
<td>f. Reading and reaction to reports</td>
<td></td>
</tr>
</tbody>
</table>

submitted by your center/project

2. To what extent do you agree with the following statements? *For each item a-d, place an X under Only One of the 4 Agreement options.*

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. NSF has been responsive in meeting my center's/project's identified needs (e.g., through telephone calls, emails).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. NSF site visits and/or evaluative actions have helped to improve the quality of my center/project.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. NSF facilitates collaboration between my center/project and other ATE projects or centers.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. NSF has an accurate understanding of my center/project.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Collaboration

Complete this section if your center/project has relationships with institutions or groups, including your center/project institutions (i.e., institutions that are the primary participants in the work of the center/project and the primary recipients of center/project funds), that provide money and/or other support.

Place an X next to

_ This Section DOES NOT APPLY
if the above paragraph does not apply to your project/center. GO TO THE NEXT SECTION OF THE SURVEY (p. 18).

Place an X next to

_ This Section DOES APPLY
if the above paragraph does describe your project/center. PROCEED WITH THE COLLABORATION SECTION.

Unless indicated otherwise, please fill out every question and items in these questions. Thank you!

All questions refer to the past 12 months.

Numerical values must be entered as integers (e.g., "3420", "6" or "0").
Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").

Collaboration with Other Institutions Or Groups

1. Non-NSF funding and in-kind support.

   I. Please CHECK FOR ALL APPLICABLE INSTITUTIONS in b-f, if you received MONETARY SUPPORT in the last 12 months (including project cost sharing). (If you place an X next to Not Applicable, please go to II. Otherwise, place an X next to all that apply).

      ___ a. This question is Not Applicable
      ___ b. Center/Project institutions
          (The institutions that are the primary participants in the work of the center/project and the primary recipients of center/project funds)
      ___ c. Business and industry
      ___ d. Public agencies (local, state, federal)
      ___ e. Educational institutions
      ___ f. Other organizations
II. Please CHECK FOR ALL APPLICABLE INSTITUTIONS in b-f, if you received IN-KIND support (non-monetary support [e.g., equipment]) in the last 12 months. (If you place an X next to Not Applicable, please go to III. Otherwise, place an X next to all that apply).

a. This question is Not Applicable
b. Center/Project institutions
   (The institutions that are the primary participants in the work of the center/project and the primary recipients of center/project funds)
c. Business and industry
d. Public agencies (local, state, federal)
e. Educational institutions
f. Other organizations

III. Total for the last 12 months. Please provide the approximate amount of total monetary and in-kind support that your center/project has received in the past 12 months from all the non-NSF sources you identified. Estimate dollar amounts to the nearest $1,000. Please enter only numbers (no "$", commas, or periods).
If the information is Unavailable enter "U".

<table>
<thead>
<tr>
<th>Description</th>
<th>Dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Total for last 12 months of monetary support</td>
<td></td>
</tr>
<tr>
<td>b. Total for last 12 months of in-kind support</td>
<td></td>
</tr>
</tbody>
</table>

2. With how many institutions EXTERNAL to your project/center has your center/project established collaborative arrangements that involve support (contributions of time, personnel sharing, equipment, etc.) and approximately how many persons from these institutions collaborate? Please specify for each type of institution listed below. For collaborators that offer their time, include only those that have spent a minimum of two days per year working with your center/project. * If the information is Unavailable enter "U"

<table>
<thead>
<tr>
<th>Types of collaborating institutions</th>
<th># of Institutions</th>
<th># of Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Business and industry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Public agencies (local, state, federal)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Educational institutions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Other organizations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. For each of the institution types 1-4, please check ALL THE COLLABORATION PURPOSES THAT APPLY for your center’s/project’s collaborative arrangements with these institutions. If you place an X next to Not Applicable, proceed to the next Institution Type. Otherwise, place an X next to all purposes that apply.

I. Institution Type 1: Business and Industry
   __ This question is Not Applicable
   __ General center or project support (e.g., advice, contributed or shared equipment/technology, contributed time and effort)
   __ Materials development (e.g., development or implementation of standards/guidelines, determining or confirming materials content, pilot testing of materials, field testing of materials)
   __ Program improvement (e.g., student recruitment program, student understanding of industry opportunities and requirements, college/school-based instruction matters, work-based instruction and experience matters, student entry to the workforce)
   __ Professional development (e.g., faculty/staff knowledge of industry needs, opportunities, and requirements; faculty/staff knowledge and skill in discipline; business/industry representatives’ knowledge of educational options and opportunities)
   __ Other. Please describe __________________________________________

II. Institution Type 2: Public Agencies (Local, State, Federal)
   __ This question is Not Applicable
   __ General center or project support (e.g., advice, contributed or shared equipment/technology, contributed time and effort)
   __ Materials development (e.g., development or implementation of standards/guidelines, determining or confirming materials content, pilot testing of materials, field testing of materials)
   __ Program improvement (e.g., student recruitment program, student understanding of industry opportunities and requirements, college/school-based instruction matters, work-based instruction and experience matters, student entry to the workforce)
   __ Professional development (e.g., faculty/staff knowledge of industry needs, opportunities, and requirements; faculty/staff knowledge and skill in discipline; business/industry representatives’ knowledge of educational options and opportunities)
   __ Other. Please describe __________________________________________
III. Institution Type 3: Educational Institutions

__ This question is Not Applicable
__ General center or project support (e.g., advice, contributed or shared equipment/technology, contributed time and effort)
__ Materials development (e.g., development or implementation of standards/guidelines, determining or confirming materials content, pilot testing of materials, field testing of materials)
__ Program improvement (e.g., student recruitment program, student understanding of industry opportunities and requirements, college/school-based instruction matters, work-based instruction and experience matters, student entry to the workforce)
__ Professional development (e.g., faculty/staff knowledge of industry needs, opportunities, and requirements; faculty/staff knowledge and skill in discipline; business/industry representatives’ knowledge of educational options and opportunities)
__ Other. Please describe ____________________________________________________________

IV. Institution Type 4: Other Organizations

__ This question is Not Applicable
__ General center or project support (e.g., advice, contributed or shared equipment/technology, contributed time and effort)
__ Materials development (e.g., development or implementation of standards/guidelines, determining or confirming materials content, pilot testing of materials, field testing of materials)
__ Program improvement (e.g., student recruitment program, student understanding of industry opportunities and requirements, college/school-based instruction matters, work-based instruction and experience matters, student entry to the workforce)
__ Professional development (e.g., faculty/staff knowledge of industry needs, opportunities, and requirements; faculty/staff knowledge and skill in discipline; business/industry representatives’ knowledge of educational options and opportunities)
__ Other. Please describe ____________________________________________________________

4. Provide ratings of the quality/productivity of collaboration by each institution type. For each item a-d, place an X under Only One of the 5 Rating options.

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Business or Industry</td>
<td></td>
</tr>
<tr>
<td>b. Public Agencies</td>
<td></td>
</tr>
<tr>
<td>(Local, State, and Federal)</td>
<td></td>
</tr>
<tr>
<td>c. Educational Institutions</td>
<td></td>
</tr>
<tr>
<td>d. Other Organizations</td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
5. Most effective collaborator

I. Which institution type has been the most effective external collaborator in helping your center/project reach its goals? Place an X next to Only One.

- Business or Industry
- Public Agencies (Local, State, and Federal)
- Educational Institutions
- Other Organizations

II. For the organization type described in 5I, briefly describe what you consider to be the two most important products and/or results of your collaboration with groups within that organization type. Optional question.

______________________________________________________

______________________________________________________

______________________________________________________

______________________________________________________
Materials Development

Complete this section if the development of materials is a focus of your center/project. "Materials" include one or more courses, modules, process models, and/or other instructional or assessment units. "Development" includes the preparation, adaptation for implementation and/or testing of materials.

**Place an X next to**

__ This Section DOES NOT APPLY

if the above paragraph does not apply to your project/center. GO TO THE NEXT SECTION OF THE SURVEY (p. 24).

**Place an X next to**

__ This Section DOES APPLY

if the above paragraph does describe your project/center. PROCEED WITH THE MATERIALS DEVELOPMENT SECTION.

Unless indicated otherwise, please fill out every question and items in these questions. Thank you! If your center/project provides instruction to students as a part of a curricular program, you should also complete a Program Improvement section.

Numerical values must be entered as integers (e.g., "3420", "6" or "0"). Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").

### Materials Development: Courses, Modules and Other Types of Materials

1. Please indicate the number of items developed or under development for each development type listed below. Materials development is often a mix of simple and substantial efforts. For example, making changes throughout a course or module would likely require substantial effort, while revision of a test would probably not require substantial effort for the center/project. List only substantial items.

   * If the information is Unavailable enter "U"

<table>
<thead>
<tr>
<th>Materials Development</th>
<th>No. in draft stage</th>
<th>No. being field tested</th>
<th>No. completed</th>
<th>No. in use locally*</th>
<th>No. In use elsewhere**</th>
<th>No. published commercially</th>
<th>No. with problem solving tasks***</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Course Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Module Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a component that can be used in more than one course)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Other. Please describe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description for c:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Materials in use locally means at institutions within your center/project.

** Materials in use elsewhere means at institutions not a part of your center/project.

*** Materials with problem solving tasks require students to 1) recognize that a problem (i.e., discrepancy between what is and what should be) exists, 2) identify possible reasons for the problem, 3) devise and implement a plan of action to resolve the problem, and/or 4) evaluate and monitor progress, revising the plan as indicated by findings.
2. Please provide the following information for up to three (3) of your best materials that your center/project developed (or is developing). You need to complete at least a-f.

a. **Material #1: Title**

b. **Material #1: Type of Development. Place an X next to Only One.**
   - Course Development
   - Module Development
   - Combination of above
   - Other. Please describe:

  c. **Material #1: Technology Field. Place an X next to Only One.**
   - Agriculture
   - Aquaculture
   - Biotechnology
   - Chemical Technology
   - Distance Learning
   - Electronics, Instrumentation, Laser and Fiber Optics
   - Engineering Technology(general)
   - Environmental Technology
   - Geographic Information Systems
   - Graphics and Multimedia Technology
   - Information Technology, Telecommunications
   - Machine Tool Technology, Metrology
   - Manufacturing and Industrial Technology
   - Marine Technology
   - Mathematics
   - Multidisciplinary or Interdisciplinary (General)
   - Physics
   - Semiconductor Manufacturing
   - Transportation
   - Other. Please describe:

  d. **Material #1: Grade Level(s). Place an X next to Only One.**
   - Elementary/Middle
   - Secondary
   - College-first year
   - College-second year
   - College-upper level
c. Material #1: Comparable Materials. Are you aware of comparable (i.e., equivalent purpose, scope, & audience) materials to Material #1 in use?  
   i. Yes  
   _ No  

   If Yes, please provide:  
   ii. Title: ________________________________________________________________  
   iii. Publisher: ____________________________________________________________  
   iv. Name of the Institution Where It Is in Use: ________________________________  

f. Material #1: Licensure or Certification Exam. Is Material #1 designed to assist students in passing a specific licensure or certification exam?  
   i. Yes  
   _ No  

   If Yes, please provide:  
   ii. Title of the Exam ________________________________________________________  

h. Material #2: Type of Development. Place an X next to Only One.  
   _ Course Development  
   _ Module Development  
   _ Combination of above  
   _ Other. Please describe: ____________________________________________________  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
i. Material #2: Technology Field. *Place an X next to Only One.*

- Agriculture
- Aquaculture
- Biotechnology
- Chemical Technology
- Distance Learning
- Electronics, Instrumentation, Laser and Fiber Optics
- Engineering Technology (general)
- Environmental Technology
- Geographic Information Systems
- Graphics and Multimedia Technology
- Information Technology, Telecommunications
- Machine Tool Technology, Metrology
- Manufacturing and Industrial Technology
- Marine Technology
- Mathematics
- Multidisciplinary or Interdisciplinary (General)
- Physics
- Semiconductor Manufacturing
- Transportation
- Other. Please describe

j. Material #2: Grade Level(s). *Place an X next to Only One.*

- Elementary/Middle
- Secondary
- College-first year
- College-second year
- College-upper level

k. Material #2: Comparable Materials. Are you aware of comparable (i.e., equivalent purpose, scope, & audience) materials to Material #2 in use?

i. __ Yes
   __ No

If Yes, please provide:

ii. Title: __________________________________________________________

iii. Publisher: _______________________________________________________

iv. Name of the Institution Where It Is in Use: __________________________
l. **Material #2: Licensure or Certification Exam.** Is Material #2 designed to assist students in passing a specific licensure or certification exam?
   i. __ Yes
      __ No

   If Yes, please provide:
   ii. Title of the Exam

m. **Material #3: Title**

n. **Material #3: Type of Development. Place an X next to Only One.**
   __ Course Development
   __ Module Development
   __ Combination of above
   __ Other. Please describe:

o. **Material #3: Technology Field. Place an X next to Only One.**
   __ Agriculture
   __ Aquaculture
   __ Biotechnology
   __ Chemical Technology
   __ Distance Learning
   __ Electronics, Instrumentation, Laser and Fiber Optics
   __ Engineering Technology (general)
   __ Environmental Technology
   __ Geographic Information Systems
   __ Graphics and Multimedia Technology
   __ Information Technology, Telecommunications
   __ Machine Tool Technology, Metrology
   __ Manufacturing and Industrial Technology
   __ Marine Technology
   __ Mathematics
   __ Multidisciplinary or Interdisciplinary (General)
   __ Physics
   __ Semiconductor Manufacturing
   __ Transportation
   __ Other. Please describe

p. **Material #3: Grade Level(s). Place an X next to Only One.**
   __ Elementary/Middle
   __ Secondary
   __ College-first year
   __ College-second year
   __ College-upper level
q. **Material #3: Comparable Materials.** Are you aware of comparable (i.e., equivalent purpose, scope, & audience) materials to Material #3 in use?
   i. _Yes_
   _No_

   If Yes, please provide:
   ii. **Title:**
   iii. **Publisher:**
   iv. **Name of the Institution Where It Is in Use:**

r. **Material #3: Licensure or Certification Exam.** Is Material #3 designed to assist students in passing a specific licensure or certification exam?
   i. _Yes_
   _No_

   If Yes, please provide:
   ii. **Title of the Exam**
3. Select one of the materials from Question 2 above as developed by your center/project. For that item please briefly describe *(Optional question):*
   a. The title of the chosen material

   ________________________________________________________________

   b. What you consider to be the most compelling evidence for its quality.

   ________________________________________________________________

   ________________________________________________________________

   ________________________________________________________________

   ________________________________________________________________

4. In the table below, identify the frequency of use for each practice that your center/project employs when developing curricular materials. *For each item a-h, place an X under Only One of the 5 Frequency options.*

<table>
<thead>
<tr>
<th>Practices</th>
<th>Frequency of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>a. Obtain verification by industry regarding alignment of materials with workforce and skill needs</td>
<td></td>
</tr>
<tr>
<td>b. Use applicable student and industry-based standards or guidelines to guide materials development</td>
<td></td>
</tr>
<tr>
<td>c. Assess student success (knowledge and skills) in comparison with standards (e.g., business/industry, educational, nontechnical skill)</td>
<td></td>
</tr>
<tr>
<td>d. Assess student success (knowledge and skills) in comparison with other nonproject or nonparticipating students</td>
<td></td>
</tr>
<tr>
<td>e. Pilot test <em>(1)</em> materials</td>
<td></td>
</tr>
<tr>
<td>f. Field-test <em>(2)</em> materials</td>
<td></td>
</tr>
<tr>
<td>internally (i.e., within the center/project)</td>
<td></td>
</tr>
<tr>
<td>g. Field-test <em>(2)</em> materials</td>
<td></td>
</tr>
<tr>
<td>externally (i.e., not center/project-based locations)</td>
<td></td>
</tr>
<tr>
<td>h. Assess improvement of student performance in the workforce</td>
<td></td>
</tr>
</tbody>
</table>

* (1) Pilot testing refers to brief, preliminary testing of materials or portions of materials; usually done with a small number of sites and/or students.

* (2) Field testing refers to testing of materials in settings where they will be used when finalized; usually large and more in-depth than pilot testing.
5. Please indicate the types of materials development in which your project engages. Place an X next to ALL THAT APPLY.

- Program improvement (e.g., developed materials used in modifying or developing courses in an ATE-funded program)
- Dissemination (e.g., commercial)
- Professional development activities
- Other. Please describe
Complete this section if your center/project provides instruction and/or support to teaching faculty and staff, so that they update their knowledge and skills in order to effectively teach new or improved curricula.

**Place an X next to**

__This Section DOES NOT APPLY__

if the above paragraph does not apply to your project/center. GO TO THE NEXT SECTION OF THE SURVEY (p. 28).

**Place an X next to**

__This Section DOES APPLY__

if the above paragraph does describe your project/center. PROCEED WITH THE PROFESSIONAL DEVELOPMENT SECTION.

Unless indicated otherwise, please fill out every question and items in these questions. Thank you! Numerical values must be entered as integers (e.g., "3420", "6" or "0"). Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").

Professional Development: Instruction and/or Support provided by your project/center to teaching faculty (e.g., college faculty and secondary school teachers).

1. Please provide the number of opportunities your project/center provided for each option for a-f (e.g., 3 conferences) and then provide the total number of participants across all opportunities for the past 12 months.

   * If the information is Unavailable enter "U"

   * If the information is Not Applicable enter “N”

<table>
<thead>
<tr>
<th>Professional Development Opportunities</th>
<th>No. of Opportunities</th>
<th>Total No. of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Conference (multiple track-participants choose from a selection of workshops or presentations to attend)</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>b. Short-term workshop (single track-1 to 3 day directed learning experience)</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>c. Inservice course or seminar (longer than a 3-day directed learning experience)</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>d. Internship, leave of absence to work with industry, or work exchange program (faculty, teachers)</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>e. On-line courses</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>f. Other (please describe)</td>
<td>_____</td>
<td>_____</td>
</tr>
</tbody>
</table>

Description for f: __________________________________________________
2. Approximately what number of participants from the following institutions was engaged in professional development opportunities provided by your center/project in the last 12 months?

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter “N”

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Secondary schools</td>
<td></td>
</tr>
<tr>
<td>b. 2-year colleges</td>
<td></td>
</tr>
<tr>
<td>c. 4-year colleges/universities</td>
<td></td>
</tr>
<tr>
<td>d. Other (please describe)</td>
<td></td>
</tr>
</tbody>
</table>

Description for d: ________________________________________________________________

3. Overall, to what extent are your professional development opportunities operating at full capacity (100 percent of available seats occupied in these opportunities)? Place an X next to Only One.

- 0-25% of full capacity
- 26-50% of full capacity
- 51-75% of full capacity
- 76-100% of full capacity

4. Follow up
I. If you formally follow up on participants in your professional development activities, please CHECK ALL FOLLOW-UP METHODS THAT APPLY. If you place an X next to Not Applicable, please go to II. Otherwise, place an X next to all that apply.

- This question is Not Applicable
- Personal (e.g., voice or in person) contacts to all participants
- Survey
- Newsletter
- Letter or email
- Other. Please describe ______________________________________________________
II. For items a-f and each of the 4 column headings (e.g., indicated satisfaction with the activity), please provide the percent of participants in the past 12 months who have taken the following actions as a result of participating in each type of professional development activity provided by your project/center.

* If the information is Unavailable enter "U"

* If the information is Not Applicable enter "N"

<table>
<thead>
<tr>
<th>Professional Development Opportunities</th>
<th>Indicated satisfaction with the activity</th>
<th>Indicated intention to use the technology, materials, and/or major ideas presented</th>
<th>Tried out the technology, materials and/or major ideas at least once in the classroom</th>
<th>Fully incorporated the technology, materials, and/or major ideas into their course or program</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Short term workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Inservice course or seminar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Internship, leave of absence to work with industry, or work exchange program</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. On-line courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Other. Please describe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description for f: ____________________________________________________________

5. Support

I. Does your center/project require participants to obtain administrative, monetary, or other support for implementation as a condition of acceptance to your professional development program?

Require Support: Place an X next to Only One.

- yes
- no

II. PLEASE CHECK ALL THAT APPLY for types of implementation support that your center/project typically provides to participants as part of your professional development program. If you place an X next to Not Applicable, please go to 6. Otherwise, place an X next to all that apply.

- This question is Not Applicable
- money
- equipment
- materials
- technical assistance
- follow-up activities (e.g., stipends, web site)
- email
- newsletter
- Other. Please describe
6. Please comment on your project's/center's effectiveness regarding professional development activities. That is, briefly describe what faculty can do now as a result of participation in professional development activities you provided that they could not do before. If possible, please provide an example. *Optional question.*

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter "N"
Program Improvement: Secondary School Level

Complete this section if your center/project provides an instructional program to students (e.g., degree, certification or other collection of courses) at the Secondary School Level and ATE grant monies have been used to improve that instructional program.

Place an X next to

This Section DOES NOT APPLY
if the above paragraph does not apply to your project/center. GO TO THE NEXT SECTION OF THE SURVEY (p. 32).

Place an X next to

This Section DOES APPLY
if the above paragraph does describe your project/center. PROCEED WITH THE PROGRAM IMPROVEMENT-SECONDARY SECTION.

Unless indicated otherwise, please fill out every question and items in these questions. Thank you! If you have modified or developed an individual course or courses in this program as part of this ATE grant, you should also complete the Materials Development section.

"Program", as used here, refers to multiple, related courses and/or field experiences for students at the designated education level. These instructional experiences lead to a defined outcome such as a degree, certification, or occupational completion point.

"Module", as used here, refers to a component that can be used in one or more courses.

"Course", as used here, refers to an educational unit (usually at the secondary, college or university level) consisting of a series of instruction periods (e.g., lectures, recitations, and laboratory sessions) dealing with a particular subject.

Numerical values must be entered as integers (e.g., "3420", "6" or "0").
Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").
Program Improvement and Student Characteristics: Secondary School Level

1. This question addresses the size and scope of your educational program(s) funded by the ATE grant for this level. For items a-d, please fill in the Total Number.

<table>
<thead>
<tr>
<th>Description</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. State the total number of ATE-grant funded programs developed/offered.</td>
<td></td>
</tr>
<tr>
<td>b. State the total number of secondary schools where the ATE-grant funded programs are offered.</td>
<td></td>
</tr>
<tr>
<td>c. State the total number of courses offered across all ATE-grant funded programs</td>
<td></td>
</tr>
<tr>
<td>d. State (estimate) the total number of students (head count) who are enrolled in one of your ATE-grant funded programs (i.e., who have taken at least one course in one of your ATE-grant funded programs during the past 12 months).</td>
<td></td>
</tr>
</tbody>
</table>

2. In completing the remainder of this section, please refer to one specific ATE-grant funded program as offered at one location and that best represents your center/project.
   a. Program name: Choose one specific ATE-grant funded program to consider when answering the remaining questions in this section.
   b. School name: Choose one location to consider when answering the remaining questions in this section.

3. Indicate the extent to which the courses in your specified ATE-grant funded program meet the following conditions. For each item a-b, place an X under Only One of the 5 column options (e.g., None).

<table>
<thead>
<tr>
<th>Condition</th>
<th>None</th>
<th>Some</th>
<th>Most</th>
<th>All</th>
<th>Don't Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Course credits can be transferred to higher degree level institutions (e.g., courses can be taken for dual credit for secondary and community college.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Certification can be obtained by students in these courses (e.g., business/industry based certification)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. How many persons instruct courses in your specified ATE-grant funded program?
5. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the Total Number of Courses and then by Course Status (New(1), Changed(2) or Unchanged(3)).

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter "N"

<table>
<thead>
<tr>
<th>Description</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Total No. of Courses in the Specified Program</td>
<td></td>
</tr>
<tr>
<td>II. Course Status</td>
<td>Number of Courses</td>
</tr>
<tr>
<td>a. New Courses(1)</td>
<td></td>
</tr>
<tr>
<td>b. Changed Courses(2)</td>
<td></td>
</tr>
<tr>
<td>c. Unchanged Courses(3)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The number of New, Changed, and Unchanged should add up to the number you entered for Total No. of Courses in the Specified Program.

* (1) New Courses means courses added as part of this grant.
* (2) Changed Courses means pre-existing courses that were substantially changed through this grant's efforts.
* (3) Unchanged Courses means pre-existing courses, used in the specified program, that were not changed through this grant's efforts.

6. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the number of your students in each of the following categories. Use the past academic year plus summer (12 months) as the basis for answering.

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter “N”

<table>
<thead>
<tr>
<th>Student Characteristics</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Number of students who applied to your specified program</td>
<td></td>
</tr>
<tr>
<td>b. Number of students enrolled in your specified program</td>
<td></td>
</tr>
<tr>
<td>c. Number of students who completed the specified program</td>
<td></td>
</tr>
<tr>
<td>d. Number of students who left the specified program without completing it</td>
<td></td>
</tr>
<tr>
<td>e. Number who go into employment as a technician</td>
<td></td>
</tr>
<tr>
<td>f. Number who continue science, technology-related, engineering, or mathematics (STEM) higher education</td>
<td></td>
</tr>
</tbody>
</table>

Of those students who completed the specified program

c. Number who go into employment as a technician

f. Number who continue science, technology-related, engineering, or mathematics (STEM) higher education
7. For courses in the single ATE-grant funded program and location you specified in Question 2, please provide your best estimate of gender, ethnicity, race, and disability information from application and enrollment information for the past academic year plus summer (12 months).

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter "N"

<table>
<thead>
<tr>
<th>a. Gender</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Male</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b. Ethnicity/Race</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(These will not necessarily sum to 100%.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Hispanic or Latino</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% American Indian or Alaska Native</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Asian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Black or African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Native Hawaiian or Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Multiracial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% White Non Hispanic/Latino</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c. Percent of students who requested accommodation due to a disability recognized under the Americans with Disabilities Act.</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students requesting ADA accommodation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NSF Award #: ____________________________

Program Improvement: Associate Degree Level (2-year college programs)

Complete this section if your center/project provides an instructional program to students (e.g. degree, certification or other collection of courses) at the Associate Degree Level (2-year college programs) and ATE grant monies have been used to improve that instructional program.

**Place an X next to**

_This Section DOES NOT APPLY_
if the above paragraph does not apply to your project/center. GO TO THE NEXT SECTION OF THE SURVEY (p. 38).

**Place an X next to**

_This Section DOES APPLY_
if the above paragraph does describe your project/center. PROCEED WITH THE PROGRAM IMPROVEMENT-ASSOCIATE SECTION.

Unless indicated otherwise, please fill out every question and items in these questions. Thank you! If you have modified or developed an individual course or courses in this program as part of this ATE grant, you should also complete the Materials Development section.

"Program", as used here, refers to multiple, related courses and/or field experiences for students at the designated education level. These instructional experiences lead to a defined outcome such as a degree, certification, or occupational completion point.

"Module", as used here, refers to a component that can be used in one or more courses.

"Course", as used here, refers to an educational unit (usually at the secondary, college or university level) consisting of a series of instruction periods (e.g., lectures, recitations, and laboratory sessions) dealing with a particular subject.

Numerical values must be entered as integers (e.g., "3420", "6" or "0"). Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").
1. This question addresses the size and scope of your ATE educational program(s) funded by the ATE grant for this level. *For items a-d, please fill in the Total Number.*

<table>
<thead>
<tr>
<th>Description</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. State the total number of ATE-grant funded programs developed/offered.</td>
<td></td>
</tr>
<tr>
<td>b. State the total number of 2-year institutions/campuses where the ATE-grant funded programs are offered.</td>
<td></td>
</tr>
<tr>
<td>c. State the total number of courses offered across all ATE-grant funded programs.</td>
<td></td>
</tr>
<tr>
<td>d. State (estimate) the total number of students (head count)</td>
<td></td>
</tr>
</tbody>
</table>

who are enrolled in one of your ATE-grant funded programs (i.e., who have taken at least one course in one of your ATE-grant funded programs during the past 12 months).

2. In completing the remainder of this section, please refer to one specific ATE-grant funded program as offered at one location and that best represents your center/project.

a. Program name: Choose one specific ATE-grant funded program to consider when answering the remaining questions in this section.

b. Institution name: Choose one institution to consider when answering the remaining questions in this section.

3. Indicate the extent to which the courses in your specified ATE-grant funded program meet the following conditions. *For each item a-c, place an X under Only One of the 5 column options (e.g., None).*

<table>
<thead>
<tr>
<th>Condition</th>
<th>None</th>
<th>Some</th>
<th>Most</th>
<th>All</th>
<th>Don't Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Course credits from secondary technical programs articulate into this program.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Course credits can be transferred to other similar institutions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Course credits can be transferred to higher degree level institutions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Which of the following options does your specified ATE-grant funded program offer (CHECK ALL THAT APPLY)? *If you place an X next to Not Applicable, please go to 5. Otherwise place an X next to all that apply.*

- This question is Not Applicable
- Multiple courses without a degree or certificate
- College certificates
- Associate degrees
- Preparation for industry-based certification

5. How many persons instruct courses in your specified ATE-grant funded program?

6. Of those persons who instruct courses in your specified ATE-grant funded program, how many also currently work in business or industry?

7. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the Total Number of Courses and then by Course Status (New(1), Changed(2) or Unchanged(3)).

*If the information is Unavailable enter "U"

*If the information is Not Applicable enter "N"

<table>
<thead>
<tr>
<th>Description</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Total No. of Courses in the Specified Program</td>
<td>_____</td>
</tr>
<tr>
<td>II. Course Status</td>
<td>Number of Courses</td>
</tr>
<tr>
<td>a. New Courses(1)</td>
<td>_____</td>
</tr>
<tr>
<td>b. Changed Courses(2)</td>
<td>_____</td>
</tr>
<tr>
<td>c. Unchanged Courses(3)</td>
<td>_____</td>
</tr>
</tbody>
</table>

**Note:** The number of New, Changed, and Unchanged should add up to the number you entered for Total No. of Courses in the Specified Program.

* (1) New Courses means courses added as part of this grant.
* (2) Changed Courses means pre-existing courses that were substantially changed through this grant's efforts.
* (3) Unchanged Courses means pre-existing courses, used in the specified program, that were not changed through this grant's efforts.
8. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the number of your students in each of the following categories. Use the past academic year plus summer (12 months) as the basis for answering.

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter "N"

<table>
<thead>
<tr>
<th>Student Characteristics</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Number of student who applied to your specified program</td>
<td></td>
</tr>
<tr>
<td>b. Number of students enrolled in your specified program</td>
<td></td>
</tr>
<tr>
<td>c. Number of students who completed the specified program</td>
<td></td>
</tr>
<tr>
<td>d. Number of students who left the specified program without completing it</td>
<td></td>
</tr>
<tr>
<td>e. Number of students who were already employed as technicians in specified program-related fields upon entry into the specified program</td>
<td></td>
</tr>
</tbody>
</table>

Of those students who completed the specified program

<table>
<thead>
<tr>
<th></th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>f. Number who go into or continue employment as a technician</td>
<td></td>
</tr>
<tr>
<td>g. Number who continue</td>
<td></td>
</tr>
<tr>
<td>science, technology-related, engineering, or mathematics (STEM) higher education</td>
<td></td>
</tr>
</tbody>
</table>

Of those students who left the specified program without completing it

<table>
<thead>
<tr>
<th></th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>h. Number who go into or continue employment as a technician</td>
<td></td>
</tr>
<tr>
<td>i. Number who continue</td>
<td></td>
</tr>
<tr>
<td>science, technology-related, engineering, or mathematics (STEM) higher education</td>
<td></td>
</tr>
</tbody>
</table>
9. For courses in the single ATE-grant funded program and location you specified in
Question 2, estimate the percent of your students in each of the following categories. Use
the past academic year plus summer (12 months) as the basis for answering.
* If the information is Unavailable enter "U"
* If the information is Not Applicable enter “N”

<table>
<thead>
<tr>
<th>Student Characteristics</th>
<th>Percent of Students (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Students who were required to take remedial math and science courses before entering your specified program</td>
<td></td>
</tr>
<tr>
<td>b. Students who meet basic science, technology, engineering, and mathematics (STEM) workforce entry requirements for technician jobs related to your specified program at the time of entry into your specified program</td>
<td></td>
</tr>
<tr>
<td>c. If your specified program offers a college certificate, what percent of the students in the specified program's courses seek this certificate?</td>
<td></td>
</tr>
<tr>
<td>d. If your specified program offers a degree, what percent of the students in the specified program's courses seek the degree?</td>
<td></td>
</tr>
</tbody>
</table>

10. For courses in the single ATE-grant funded program and location you specified in
Question 2, please provide your best estimate of gender, ethnicity, race, and disability
information from application and enrollment information for your ATE grant-based
academic specified program for the past academic year plus summer (12 months).
* If the information is Unavailable enter "U"
* If the information is Not Applicable enter “N”

<table>
<thead>
<tr>
<th>a. Gender</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Male</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b. Ethnicity/Race</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Hispanic or Latino</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% American Indian or Alaska Native</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Asian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Black or African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Native Hawaiian or Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Multiracial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% White Non Hispanic/Latino</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c. Percent of students who requested accommodation due to a disability recognized under the Americans with Disabilities Act.</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
</table>
Program Improvement: Baccalaureate Level (4-year college/university programs)

Complete this section if your center/project provides an instructional program to students (e.g., degree, certification or other collection of courses) at the Baccalaureate Level (4-year college/university programs) and ATE grant monies have been used to improve that instructional program.

Place an X next to
__ This Section DOES NOT APPLY
if the above paragraph does not apply to your project/center. YOU HAVE NOW COMPLETED THE SURVEY. PLEASE GO TO THE FILLING OUT INSTRUCTIONS FOR INFORMATION ON WHERE TO SEND IT.

Place an X next to
__ This Section DOES APPLY
if the above paragraph does describe your project/center. PROCEED WITH THE PROGRAM IMPROVEMENT-BACCALAUREATE SECTION.

Unless indicated otherwise, please fill out every question and items in these questions. Thank you! If you have modified or developed an individual course or courses in this program as part of this ATE grant, you should also complete the Materials Development section.

"Program", as used here, refers to multiple, related courses and/or field experiences for students at the designated education level. These instructional experiences lead to a defined outcome such as a degree, certification, or occupational completion point.

"Module", as used here, refers to a component that can be used in one or more courses.

"Course", as used here, refers to an educational unit (usually at the secondary, college or university level) consisting of a series of instruction periods (e.g., lectures, recitations, and laboratory sessions) dealing with a particular subject.

Numerical values must be entered as integers (e.g., "3420", "6" or "0"). Do not use: decimal points, dollar signs, commas or percent signs in numerical values (e.g., "3,000", "6.00", "$320" or "95").
Program Improvement and Student Characteristics: Baccalaureate Level (4-year college/university programs)

1. This question addresses the size and scope of your educational program(s) funded by the ATE grant for this level. For items a-d, please fill in the Total Number.

   **Description**
   a. State the total number of ATE-grant funded programs developed/offered.
   b. State the total number of 4-year institutions/campuses where the ATE-grant funded programs are offered.
   c. State the total number of courses offered across all ATE-grant funded programs.
   d. State (estimate) the total number of students (head count) who are enrolled in one of your ATE-grant funded programs (i.e., who have taken at least one course in one of your ATE-grant funded programs during the past 12 months).

2. In completing the remainder of this section, please refer to one specific ATE-grant funded program as offered at one location and that best represents your center/project.
   a. Program name: Choose one specific ATE-grant funded program to consider when answering the remaining questions in this section.
   b. Institution name: Choose one location to consider when answering the remaining questions in this section.

3. Indicate the extent to which the courses in your specified ATE-grant funded program meet the following conditions. For each item a-f, place an X under Only One of the 5 column options (e.g., None).

   **Condition**
   a. Course credits from secondary technical programs articulate into this program.
   b. Course credits from associate degree technical programs articulate into this program.
   c. Course credits can be transferred to other similar institutions.
   d. Course credits can be transferred to higher degree level institutions.
   e. Preparation for industry-based certification.
   f. Provides a baccalaureate degree in a technician-based program.
4. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the Total Number of Courses and then by Course Status (New(1), Changed(2) or Unchanged(3)).

* If the information is Unavailable enter "U"

<table>
<thead>
<tr>
<th>Description</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Total No. of Courses in the Specified Program</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td>Number of Courses</td>
</tr>
<tr>
<td>Course Status</td>
<td></td>
</tr>
<tr>
<td>a. New Courses(1)</td>
<td></td>
</tr>
<tr>
<td>b. Changed Courses(2)</td>
<td></td>
</tr>
<tr>
<td>c. Unchanged Courses(3)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The number of New, Changed, and Unchanged should add up to the number you entered for Total No. of Courses in the Specified Program.

* (1) New Courses means courses added as part of this grant.
* (2) Changed Courses means pre-existing courses that were substantially changed through this grant's efforts.
* (3) Unchanged Courses means pre-existing courses, used in the specified program, that were not changed through this grant's efforts.

5. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the number of your students in each of the following categories. (Use the past academic year plus summer (12 months) as the basis for answering.)

* If the information is Unavailable enter "U"

<table>
<thead>
<tr>
<th>Student Characteristics</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Number of student who applied to your specified program</td>
<td></td>
</tr>
<tr>
<td>b. Number of students enrolled in your specified program</td>
<td></td>
</tr>
<tr>
<td>c. Number of students who completed the specified program without completing it</td>
<td></td>
</tr>
<tr>
<td>d. Number of students who left the specified program</td>
<td></td>
</tr>
<tr>
<td>e. Number of students who were already employed as technicians in specified program-related fields upon entry into the specified program</td>
<td></td>
</tr>
</tbody>
</table>
Of those students who completed the specified program

f. Number who go into or continue employment as a technician

g. Number who continue
   science, technology-related, engineering, or mathematics (STEM) higher education

Of those students who left the specified program without completing it

h. Number who go into or continue employment as a technician

i. Number who continue
   science, technology-related, engineering, or mathematics (STEM) higher education

6. For courses in the single ATE-grant funded program and location you specified in Question 2, estimate the percent of your students in each of the following categories. Use the past academic year plus summer (12 months) as the basis for answering.
* If the information is Unavailable enter "U"
* If the information is Not Applicable enter "N"

Student Characteristics

<table>
<thead>
<tr>
<th>Category</th>
<th>Percent of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Students who were required to take remedial science and math courses before entering your specified program</td>
<td></td>
</tr>
<tr>
<td>b. Students who meet basic science, technology, engineering, and mathematics (STEM) workforce entry requirements for technician jobs related to your specified program at the time of entry into your specified program</td>
<td></td>
</tr>
<tr>
<td>c. If your specified program offers a certificate, what percent of the students in the specified program's courses seek this certificate?</td>
<td></td>
</tr>
<tr>
<td>d. If your specified program offers a degree, what percent of the students in the specified program's courses seek the degree?</td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
7. For courses in the single ATE-grant funded program and location you specified in Question 2, please provide your best estimate of gender, ethnicity, race, and disability information from application and enrollment information for the past academic year plus summer (12 months).

* If the information is Unavailable enter "U"
* If the information is Not Applicable enter “N”

<table>
<thead>
<tr>
<th>a. Gender</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Male</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. Ethnicity/Race

(These will not necessarily sum to 100%.)

<table>
<thead>
<tr>
<th>Student Category</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Hispanic or Latino</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% American Indian or Alaska Native</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Asian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Black or African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Native Hawaiian or Other Pacific Islander</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Multiracial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% White Non Hispanic/Latino</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c. Percent of students who requested accommodation due to a disability recognized under the Americans with Disabilities Act.

<table>
<thead>
<tr>
<th>Student requesting ADA accommodation</th>
<th>Applicants (%)</th>
<th>Enrollment (%)</th>
</tr>
</thead>
</table>
APPENDIX E

Measures Detail
Key

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF</td>
<td>Indicates the actual amount of the grant used by a project at the time it responded to the survey</td>
</tr>
<tr>
<td></td>
<td>Basic information</td>
</tr>
<tr>
<td></td>
<td>1d Funding start date</td>
</tr>
<tr>
<td></td>
<td>1e Funding end date</td>
</tr>
<tr>
<td></td>
<td>1f Total award amount</td>
</tr>
<tr>
<td></td>
<td>Length = (end date - start date)/365</td>
</tr>
<tr>
<td></td>
<td>Age = (survey date - start date)/365</td>
</tr>
<tr>
<td></td>
<td>NSF = ([total award amount]/[length])*[age]</td>
</tr>
<tr>
<td>AGE</td>
<td>The number of years the project has been operating at the time it responded to the survey</td>
</tr>
<tr>
<td></td>
<td>Basic information</td>
</tr>
<tr>
<td></td>
<td>1d Funding start date</td>
</tr>
<tr>
<td></td>
<td>AGE = (survey date - start date)/365</td>
</tr>
</tbody>
</table>
### Evaluative Practices (EVAL)

The proportion of years where a project engaged an internal, external, or both types of evaluators.

**PI Overview**
3a type of evaluator hired

1 = no evaluator; 2 = external; 3 = internal; 4 = both types

Data recoded so that 0 = no evaluator & 1 = any type (2, 3, or 4)

For projects with only one response, values were 0 or 1. For projects with multiple annual responses, value was average of annual responses, representing a proportion of years when an evaluator was used.

### Monitoring (MNTR)

The number of different NSF monitoring activities in which a project participated in a given year.

**Monitoring**
1. NSF monitoring of centers and projects
   a. Site visits [by NSF]
   b. Telephone calls
   c. Email contacts
   d. Visits to NSF
   e. Principal Investigator meetings
   f. Reading and reaction to reports submitted by your center/project

1 = no times; 2 = 1 time; 3 = 2-4 times; 4 = more than 4 times

Data recoded so that 0 = no times and 1 = at least one time (2, 3, or 4).

\[ MNTR = \text{sum (1a..1f)} \]
## Collaboration Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
</table>
| **EXT_SUP** | The amount of combined monetary and in-kind support received by a project in the last 12 months.  
Collaboration  
1. III a Total monetary support from non-NSF sources  
1. III b Total in-kind support from non-NSF sources  
EXT_SUP = IIIa + IIIb |
| **PTNRS** | The total number of institutions with which the project collaborated.  
Collaboration  
2. Number of institutions external to your organization with which you have collaborative arrangements  
a. Business and industry  
b. Public institutions  
c. Educational institutions  
d. Other types  
PTNRS = sum (2a..2d) |
### Materials Development Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRSE</td>
<td>A score reflecting development of course materials based on the number of courses in various stages of development.</td>
</tr>
<tr>
<td></td>
<td><strong>Materials Development</strong></td>
</tr>
<tr>
<td></td>
<td>1a. Number of courses in various stages of development</td>
</tr>
<tr>
<td></td>
<td>Stages are—(i) draft, (ii) field tested, (iii) completed, (iv) in use locally, (v) in use elsewhere, and (vi) published commercially.</td>
</tr>
<tr>
<td></td>
<td>For each course in each category, 1, 2, 4, 5, 7, 10 points, respectively, were awarded.</td>
</tr>
<tr>
<td></td>
<td>$1\text{ai} = 1\text{ai} \times 1; 1\text{aii} = 1\text{aii} \times 2; 1\text{aiii} = 1\text{aiii} \times 4; 1\text{aiv} = 1\text{aiv} \times 5; 1\text{av} = 1\text{av} \times 7; 1\text{avi} = 1\text{avi} \times 10$</td>
</tr>
<tr>
<td></td>
<td>CRSE = $1\text{ai} + 1\text{aii} + 1\text{aiii} + 1\text{aiv} + 1\text{av} + 1\text{avi}$</td>
</tr>
<tr>
<td>MDLE</td>
<td>A score reflecting development of modules based on the number of modules in various stages of development.</td>
</tr>
<tr>
<td></td>
<td><strong>Materials Development</strong></td>
</tr>
<tr>
<td></td>
<td>1b. Number of modules in various stages of development</td>
</tr>
<tr>
<td></td>
<td>Stages are—(i) draft, (ii) field tested, (iii) completed, (iv) in use locally, (v) in use elsewhere, and (vi) published commercially.</td>
</tr>
<tr>
<td></td>
<td>For each course in each category, 1, 2, 4, 5, 7, 10 points, respectively, were awarded.</td>
</tr>
<tr>
<td></td>
<td>$1\text{bi} = 1\text{bi} \times 1; 1\text{bii} = 1\text{bii} \times 2; 1\text{biii} = 1\text{biii} \times 4; 1\text{biv} = 1\text{biv} \times 5; 1\text{bv} = 1\text{bv} \times 7; 1\text{bvi} = 1\text{bvi} \times 10$</td>
</tr>
<tr>
<td></td>
<td>MDLE = $(1\text{bi} + 1\text{bii} + 1\text{biii} + 1\text{biv} + 1\text{bv} + 1\text{bvi}) \times .5$</td>
</tr>
<tr>
<td>OTH_MAT</td>
<td>A score reflecting development of other types of materials based on the number of courses in various stages of development.</td>
</tr>
<tr>
<td></td>
<td><strong>Materials Development</strong></td>
</tr>
<tr>
<td></td>
<td>1c. Number of other materials in various stages of development</td>
</tr>
<tr>
<td></td>
<td>Stages are—(i) draft, (ii) field tested, (iii) completed, (iv) in use locally, (v) in use elsewhere, and (vi) published commercially.</td>
</tr>
<tr>
<td></td>
<td>For each course in each category, 1, 2, 4, 5, 7, 10 points, respectively, were awarded.</td>
</tr>
<tr>
<td></td>
<td>$1\text{ci} = 1\text{ci} \times 1; 1\text{cii} = 1\text{cii} \times 2; 1\text{ciii} = 1\text{ciii} \times 4; 1\text{civ} = 1\text{civ} \times 5; 1\text{cv} = 1\text{cv} \times 7; 1\text{cvi} = 1\text{cvi} \times 10$</td>
</tr>
<tr>
<td></td>
<td>OTH_MAT = $(1\text{ci} + 1\text{cii} + 1\text{ciii} + 1\text{civ} + 1\text{cv} + 1\text{cvi}) \times .25$</td>
</tr>
</tbody>
</table>
Professional Development Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPPORT</td>
<td>A score reflecting numbers and types of professional development opportunities offered.</td>
</tr>
</tbody>
</table>

**Professional Development**
1. Number of professional development opportunities offered
   a. Conference
   b. Short-term workshop
   c. Inservice course or seminar
   d. Internship, leave of absence to work with industry, or work exchange program
   e. On-line courses
   f. Other

For each type of opportunity, 2, 3, 5, 5, 3, and 1 points were awarded

\[
1_{ai} = 1_{ai} \times 2; \quad 1_{bi} = 1_{bi} \times 3; \quad 1_{ci} = 1_{ci} \times 5; \quad 1_{di} = 1_{di} \times 5; \quad 1_{ei} = 1_{ei} \times 3; \quad 1_{fi} = 1_{fi}
\]

**OPPORT** = sum \((1_{ai}, 1_{bi}, \ldots, 1_{fi})\)

**PARTS** Score reflecting the total number of participants across all types of professional development opportunities offered.

**Professional Development**
1. Number of participants across all opportunities in each type offered
   a. Conference (multiple track-participants choose from a selection of workshops or presentations to attend)
   b. Short-term workshop (single track-1 to 3 day directed learning experience)
   c. Inservice course or seminar (longer than a 3-day directed learning experience)
   d. Internship, leave of absence to work with industry, or work exchange program (faculty, teachers)
   e. On-line courses
   f. Other

**PARTS** = sum \((1_{a1}, 1_{b1}, \ldots, 1_{f1})\)
Program Improvement Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRGMS</strong></td>
<td>The total number of grant-funded programs offered.</td>
</tr>
<tr>
<td>Program improvement—secondary</td>
<td></td>
</tr>
<tr>
<td>Program improvement—associate</td>
<td></td>
</tr>
<tr>
<td>Program improvement—baccalaureate</td>
<td></td>
</tr>
<tr>
<td>1a. The total number of ATE-grant funded programs developed/offered.</td>
<td></td>
</tr>
<tr>
<td><strong>PRGMS = 1a—secondary + 1a—associate + 1a—baccalaureate</strong></td>
<td></td>
</tr>
<tr>
<td><strong>SCHLS</strong></td>
<td>The total number of schools where the grant-funded programs are offered.</td>
</tr>
<tr>
<td>Program improvement—secondary</td>
<td></td>
</tr>
<tr>
<td>Program improvement—associate</td>
<td></td>
</tr>
<tr>
<td>Program improvement—baccalaureate</td>
<td></td>
</tr>
<tr>
<td>1b. The total number of schools were ATE-grant funded programs are offered.</td>
<td></td>
</tr>
<tr>
<td><strong>SCHLS = 1b—secondary + 1b—associate + 1b—baccalaureate</strong></td>
<td></td>
</tr>
<tr>
<td><strong>CRSES</strong></td>
<td>The total courses offered across all of the programs.</td>
</tr>
<tr>
<td>Program improvement—secondary</td>
<td></td>
</tr>
<tr>
<td>Program improvement—associate</td>
<td></td>
</tr>
<tr>
<td>Program improvement—baccalaureate</td>
<td></td>
</tr>
<tr>
<td>1c. The total number of courses offered across all of the ATE-grant funded programs.</td>
<td></td>
</tr>
<tr>
<td><strong>CRSES = 1c—secondary + 1c—associate + 1c—baccalaureate</strong></td>
<td></td>
</tr>
</tbody>
</table>
### Impact Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
</table>
| **STDNTS** | The total number of students who enrolled in at least one course in one ATE-funded program  
Program improvement—secondary  
Program improvement—associate  
Program improvement—baccalaureate  
1d. The total number of students who enrolled in at least one course in one ATE-grant funded program.  

\[ STDNTS = 1d{\text{-secondary}} + 1d{\text{-associate}} + 1d{\text{-baccalaureate}} \] |
| **CMPLT** | The total number of students who completed a specified program  
Program improvement—secondary  
6c. Number of students who completed the specified program  
Program improvement—associate  
8c. Number of students who completed the specified program  
Program improvement—baccalaureate  
5c. Number of students who completed the specified program  

\[ CMPLT = 6c{\text{-secondary}} + 8c{\text{-associate}} + 5c{\text{-baccalaureate}} \] |
| **PLCMNT** | The number of students who completed a specified program who are placed, or continue working, in a related technician field  
Program improvement—secondary  
6e. Number of students who continue or start work  
Program improvement—associate  
8g. Number of students who continue or start work  
Program improvement—baccalaureate  
5f. Number of students who continue or start work  

\[ PLCMNT = 6e{\text{-secondary}} + 8g{\text{-associate}} + 5f{\text{-baccalaureate}} \] |
### Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>AGE</th>
<th>NSF</th>
<th>EVAL</th>
<th>MNTR</th>
<th>EXTSUP</th>
<th>STNRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td>-0.129</td>
<td>0.411</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVAL</td>
<td>0.152</td>
<td>-0.075</td>
<td>0.300</td>
<td>1.000</td>
<td>0.086</td>
<td>0.157</td>
</tr>
<tr>
<td>MNTR</td>
<td>0.300</td>
<td>1.000</td>
<td>0.086</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>EXTSUP</td>
<td>0.157</td>
<td>0.157</td>
<td>1.000</td>
<td>0.102</td>
<td>0.039</td>
<td>0.168</td>
</tr>
<tr>
<td>STNRS</td>
<td>0.102</td>
<td>0.039</td>
<td>0.168</td>
<td>1.000</td>
<td>0.168</td>
<td>0.222</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CRSE</th>
<th>MDELE</th>
<th>OTH_MAT</th>
<th>OPPORT</th>
<th>PART</th>
<th>PRGRMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRSE</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDELE</td>
<td>0.074</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTH_MAT</td>
<td>-0.075</td>
<td>0.134</td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPPORT</td>
<td>0.286</td>
<td>0.106</td>
<td>0.216</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PART</td>
<td>0.263</td>
<td>0.183</td>
<td>0.208</td>
<td>0.661</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PRGRMS</td>
<td>0.359</td>
<td>0.162</td>
<td>0.037</td>
<td>0.258</td>
<td>0.310</td>
<td>1.000</td>
</tr>
<tr>
<td>SCHLS</td>
<td>0.366</td>
<td>0.106</td>
<td>0.105</td>
<td>0.348</td>
<td>0.448</td>
<td>0.600</td>
</tr>
<tr>
<td>CRSES</td>
<td>0.442</td>
<td>0.146</td>
<td>0.000</td>
<td>0.440</td>
<td>0.407</td>
<td>0.658</td>
</tr>
<tr>
<td>STNNTS</td>
<td>0.437</td>
<td>0.104</td>
<td>-0.045</td>
<td>0.390</td>
<td>0.422</td>
<td>0.630</td>
</tr>
<tr>
<td>CMPLTS</td>
<td>0.140</td>
<td>0.236</td>
<td>0.031</td>
<td>0.189</td>
<td>0.200</td>
<td>0.425</td>
</tr>
<tr>
<td>PLCMNT</td>
<td>0.038</td>
<td>0.238</td>
<td>0.053</td>
<td>0.237</td>
<td>0.253</td>
<td>0.319</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SCHLS</th>
<th>CRSES</th>
<th>STNNTS</th>
<th>CMPLTS</th>
<th>PLCMNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHLS</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRSES</td>
<td>0.543</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STNNTS</td>
<td>0.567</td>
<td>0.696</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMPLTS</td>
<td>0.350</td>
<td>0.259</td>
<td>0.354</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>PLCMNT</td>
<td>0.364</td>
<td>0.312</td>
<td>0.403</td>
<td>0.583</td>
<td>1.000</td>
</tr>
</tbody>
</table>
APPENDIX G

PRELIS and LISREL Code
Data screening for original data set

!Screen Original Dataset
SY='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\SEMdatal.PSF'
CO ALL
OU MA=CM AC=SEMdata1.acm

Data screening for standardized data set

!Screen z-score dataset
SY='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\SEMdataz.PSF'
CO ALL
OU MA=CM AC=SEMdataz.acm

Normalize z-score dataset

!Normalize z-score dataset
SY='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\SEMdataz.PSF'
CO ALL
NS ALL
OU MA=CM RA=SEMDATAznor.psf AC=SEMDATAznor.acm PA

Calculation of Covariance Matrix

!Covariance Matrix for znormal
SY='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\SEMDATAznor.PSF'
OU MA=CM
Full Model

TI ATE Full Model 2
DA NI=17 NO=115 NG=1 MA=CM
RA F1="C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.psf"
AC="C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.acm"
SE
5 6 7 8 9 10 11 12 13 14 15 16 17 1 2 3 4 /
MO NX=4 NY=13 NK=2 NE=5 LY=FU,FI LX=FU,FI BE=FU,FI GA=FU,FI PH=SY,FR
PS=D1,FR TE=SY TD=SY
LE
collab matdev profdev progimp impact
LK
projchar orgprac
FR LY(1,1) LY(2,1) LY(3,2) LY(4,2) LY(5,2) LY(6,3) LY(7,3) LY(8,4) LY(9,4)
FR LY(10,4) LY(11,5) LY(12,5) LY(13,5) LX(1,1) LX(2,1) LX(3,2) LX(4,2) BE(2,1)
FR BE(3,1) BE(3,2) BE(4,1) BE(4,2) BE(4,3) BE(5,4) GA(1,1) GA(1,2)
FR TE(13,12)
PD
OU ME=ML AD=OFF IT=500 MI RS

Collaboration-Impact Model

TI CI Model
DA NI=17 NO=115 NG=1 MA=CM
RA F1="C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.psf"
AC="C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.acm"
SE
7 8 10 11 12 13 14 15 16 5 6 /
MO NX=2 NY=9 NK=1 NE=4 LY=FU,FI LX=FU,FI BE=FU,FI GA=FU,FI PH=SY,FR
PS=D1,FR TE=SY TD=SY
LE
matdev profdev progimp impact
LK
collab
FR LY(1,1) LY(2,1) LY(3,2) LY(4,2) LY(5,3) LY(6,3) LY(7,3) LY(8,4) LY(9,4)
FR LX(1,1) LX(2,1)
FR BE(2,1) BE(3,1) BE(3,2) BE(4,3) GA(1,1) GA(2,1) GA(3,1)
PD
OU ME=ML AD=OFF IT=1000 MI RS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Collaboration-Impact Model no Materials Development

TI CI Model no MD
DA NI=17 NO=115 NG=1 MA=CM
RA FI='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.psf'
AC='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.acm'
SE 10 11 12 13 14 15 16 5 6 /
MO NX=2 NY=7 NK=1 NE=3 LY=FU,FI LX=FU,FI BE=FU,FI GA=FU,FI PH=SY,FR
PS=DI,FR TE=SY TD=SY
LE profdev progimp impact
LK
collab
FR LY(1,1) LY(2,1) LY(3,2) LY(4,2) LY(5,2) LY(6,3) LY(7,3)
FR LX(1,1) LX(2,1)
FR BE(2,1) BE(3,2) GA(1,1) GA(2,1)
PD
OU ME=ML AD=OFF IT=1000 MI RS

Full Model no Materials Development

TI ATE Full Model no MD 2
DA NI=17 NO=115 NG=1 MA=CM
RA FI='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.psf'
AC='C:\Documents and Settings\Carl Hanssen\My Documents\PhD\Dissertation\Data\semdataznor.acm'
SE 5 6 10 11 12 13 14 15 16 17 1 2 3 4 /
MO NX=4 NY=10 NK=2 NE=4 LY=FU,FI LX=FU,FI BE=FU,FI GA=FU,FI PH=SY,FR
PS=DI,FR TE=SY TD=SY
LE
collab profdev progimp impact
LK
projchar orgprac
FR LY(1,1) LY(2,1) LY(3,2) LY(4,2) LY(5,3) LY(6,3) LY(7,3) LY(8,4) LY(9,4)
FR LY(10,4) LX(1,1) LX(2,1) LX(3,2) LX(4,2) BE(2,1)
FR BE(3,1) BE(3,2) BE(4,3) GA(1,1) GA(1,2)
FR TE(10,9)
PD
OU ME=ML AD=OFF IT=500 MI RS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.