

SPRINGER, ROBERT, Ph.D. Applying the Use, Methods, and Values Branches of Evaluations' Theory Tree to Institutional Research. (2011)  
Directed by Dr. John Willse. 182 pp.

Evaluation and Institutional Research (IR) both share similar beginnings, purpose, reactions, and concerns. The primary difference between the two professions is that Evaluation is an established discipline, whereas IR appears to have only started moving in that direction. This study mapped an Evaluation Theory Tree containing three branches (Values, Methods, and Use) to Institutional Research. A national survey collected practicing IR professionals' responses to questions that related to stakeholder and decision maker involvement in studies, report/study processes and procedures, and directors' approaches to conducting IR.

Confirmatory factor analysis results provided confirmation of that three-branch tree structure. Multivariate analysis of variance (MANOVA) results indicated group differences between Evaluators and IRs with Evaluators scoring higher in several areas of professional practice. Tests results were compelling. Findings indicated that there are differences between Evaluators and IR in their reported practices and procedures—particularly in the area of assessment/evaluation. Multiple linear regression (MLR) was used to evaluate approaches in conducting IR and employing best practices in reporting processes and procedures. Results provide implications for conceptualizing connections among the Evaluation Theory Tree branches. Lastly, relationships

between office staffing and institutional type with participants responses to various questions was also tested.

Overall, this study's results indicate that there are potential benefits to employing Evaluation theory to IR. For examples, evaluation processes and procedures could help guide IR practice in research and reporting approaches, employing techniques that enhance utilization of reports and studies, and enhancing stakeholder involvement in the collection and analysis of data.

APPLYING THE USE, METHODS, AND VALUES BRANCHES  
OF EVALUATIONS' THEORY TREE TO  
INSTITUTIONAL RESEARCH

by

Robert Springer

A Dissertation Submitted to  
the Faculty of The Graduate School at  
The University of North Carolina at Greensboro  
in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Greensboro  
2011

Approved by

Dr. John Willse  
Committee Chair

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair Dr. John Willse

Committee Members Dr. Terry Ackerman

Dr. Bob Henson

Dr. Sarah Carrigan

March 11, 2011  
Date of Acceptance by Committee

March 11, 2011  
Date of Final Oral Examination

## ACKNOWLEDGMENTS

There are a number of people who have been instrumental in their support, guidance, and encouragement with me in this dissertation journey. Deb Bartz was clearly instrumental in helping shape this study. Her seminar class on Evaluation was informative, enlightening, and engaging—thank you for sharing your knowledge and experiences. Dr. John Willse, my committee chair, deserves a debt of gratitude. His clear and concise guidance made completing this dissertation possible. Feedback from my committee helped to strengthen the conceptual framework for the study (Dr. Terry Ackerman, Dr. Bob Henson, Dr. Sarah Carrigan, and Dr. John Willse). Lastly, my family has been supportive, encouraging, and motivating during this entire time. Maggie and Grant, my children, also deserve recognition—thank you for your patience. My wife (Becky) has since the day I met her been instrumental in my life. She is my partner—in every sense of the word. Her support has been immeasurable and has allowed me to spend countless weekends and week nights completing this journey.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	vi
LIST OF FIGURES .....	viii
CHAPTER	
I. INTRODUCTION.....	1
Statement of the Problem .....	2
Purpose of the Study and Research Questions .....	3
II. LITERATURE REVIEW.....	5
Overall Purpose .....	5
History .....	10
Educational Backgrounds .....	18
Contextually Driven .....	19
Method Proclivity.....	28
Stakeholder Involvement.....	31
Internal-External Role .....	36
Summative-Formative .....	42
Utilization Concerns .....	44
Discipline.....	55
Theory.....	57
Similarities: A Final Word .....	58
A Possible Theory: The Evaluation Theory Tree.....	60
Methods Branch .....	61
Values Branch .....	62
Use Branch.....	63
Low Hanging Fruit .....	64
Reality .....	69
Summary.....	70
III. METHODS .....	73
Survey Administration .....	73
Theory Fit: Confirmatory Factor Analysis .....	79
Branch Placement: Multivariate Analysis of Variance and Descriptive Discriminant Analysis .....	80

Approaches Branches and Core Reporting: Multiple Linear Regression .....	82
Relationship between Branch Score and Office FTE: Scatter Diagrams .....	83
Relationship between Branch Score and School Type: Box Plots .....	84
IV. RESULTS.....	85
Demographic Information.....	85
Confirmatory Factor Analysis Results .....	93
MANOVA with Discriminant Follow Up Results.....	105
Discriminant Function Analysis .....	110
Multiple Linear Regression Results.....	112
Branch Scores: Function of FTE .....	120
Branch Scores: Function of School Type .....	121
V. CONCLUSIONS AND DISCUSSION .....	125
Limitations of this Study .....	130
Future Research .....	131
Conclusions .....	132
REFERENCES .....	134
APPENDIX A. SURVEY INSTRUMENT .....	153
APPENDIX B. Q-Q PLOTS .....	165
APPENDIX C. SCATTER PLOTS.....	175
APPENDIX D. BOX PLOTS.....	179

## LIST OF TABLES

	Page
Table 1. Demographic Description of Participants .....	86
Table 2. Work Experience and Knowledge .....	87
Table 3. Cross Tabulation: Director of IR and Years in IR .....	88
Table 4. Highest Degree of Respondents .....	89
Table 5. CORE IR Reporting Descriptive Statistics .....	90
Table 6. Assessment/Evaluation Practices Descriptive Statistics .....	91
Table 7. Approaches to Conducting IR Descriptive Statistics .....	92
Table 8. Coefficient Alpha .....	93
Table 9. CORE IR Reporting Model Results .....	94
Table 10. Estimates, Standardized Estimates, and Standard Errors for the 3-Branch Model .....	95
Table 11. Covariance among Branches (Factors) .....	95
Table 12. Assessment/Evaluation Model Results .....	97
Table 13. Estimates, Standardized Estimates, and Standard Errors for Each Item in the 3-Branch Model .....	98
Table 14. Covariance among Branches (Factors) .....	99
Table 15. Approaches to Conducting IR Model Results .....	100
Table 16. Estimates, Standardized Estimates, and Standard Errors for Each Item in the 3-Branch Model .....	101
Table 17. Covariance among Branches (Factors) .....	102
Table 18. Pearson Correlations .....	104

Table 19.	Mean Scores by Branch for Evaluator/IR.....	105
Table 20.	Box's Test of Equality of Covariance Matrices.....	106
Table 21.	MANOVA Results for Eval/IR.....	107
Table 22.	Levene's Test of Equality of Error Variances.....	108
Table 23.	Tests of Between-Subjects Effects.....	109
Table 24.	Eigenvalues.....	110
Table 25.	Structure Matrix.....	111
Table 26.	Mean CORE IR Reporting Values Branch: Approaches to IR Items Coefficients.....	113
Table 27.	Mean CORE IR Reporting Methods Branch: Approaches to IR Items Coefficients.....	115
Table 28.	Mean CORE IR Reporting Use Branch: Approaches to IR Items Coefficients.....	118
Table 29.	Descriptive Information.....	122
Table 30.	Test of Homogeneity of Variances.....	123
Table 31.	One-Way ANOVA Results.....	123
Table 32.	Multiple Comparisons.....	124

## LIST OF FIGURES

	Page
Figure 1. Path Diagram for the 3-Branch Model (CORE IR Reporting) .....	96
Figure 2. Path Diagram for the 3-Branch Model (Assessment/Evaluation Practices) .....	99
Figure 3. Path Diagram for the 3-Branch Model (Approaches to Conducting IR) .....	103
Figure 4. Representative Scatterplot: FTE x CORE IR Reporting Methods Branch.....	120
Figure 5. CORE IR Reporting Use Branch Box Plot .....	122

## CHAPTER I

### INTRODUCTION

As professions, do Institutional Research (IR) and evaluation have more in common than we think? While both professions appear to have strong similarities, there are differences. One distinction is clear; evaluation has evolved more as a discipline than institutional research. As a result, IR might find that evaluation has much to offer in its own development. This statement does not imply that IR is not a profession. To the contrary, IR is a well established profession. But, it does have room for growth and taking some lessons from evaluation's experience might shed some light on how to best manage its own progression.

What are the commonalities between institutional research (IR) and evaluation? What are their differences? What can each learn from each other? While there appears to be strong similarities between these professions, direct comparisons between the two professions are limited. Still with such noticeable similarities and evaluation as an established discipline, it is surprising that IR does not reference evaluation methods or approaches. For instance, both professions employ concepts, methods, and theories adapted from other disciplines. In particular, both professions tend to reference the disciplines of economics, psychology, and sociology in their respective research. One basis for

the lack of referencing might be that IR is simply unaware of the potential benefits from evaluation. After all, evaluation is a relatively new discipline and in an age of abundant data, it takes time to filter new and useful information in meaningful ways. If each profession is shown to have enough similarities, it might be possible to draw inferences from one profession to assist the other.

Much of the published research tends to be associated with their respective journals. For example, Institutional Research (IR) related articles tend to be found in either *New Directions for Institutional Research* or *Research in Higher Education* while Evaluation related articles tend to be nested within *New Directions for Evaluation*, *American Journal of Evaluation*, and *Evaluation Practice*. In addition, each profession's international membership organization produces publications that provide valuable information. These publications tend to support the professional development of the organization and provide a great source for novice and experienced researchers alike. Finally, because IR is not a discipline, it should not be surprising that IR information does not reside in text books at the same level of Evaluation; whereas additional information about Evaluation, which is a discipline, can be found in various text books and books.

### **Statement of the Problem**

Degree programs typically have a set of theories that undergird the discipline. While it makes sense that theory is developed after the discipline is created rather than vice-versa, it appears that having a theory based program is a typical hallmark for being recognized as a discipline. IR appears to be no

different in following a logical path towards becoming a discipline. For example, post-graduate IR certificates are available at a number of schools and at least one school (the University of North Carolina at Greensboro) is proposing a graduate degree in Institutional Research. However and to date, there is no recognized theory of institutional research. An IR theory is needed for it to be recognized as a discipline.

Why is developing an IR theory important? The most direct answer is that theory helps to guide practice. IR practitioners have formal training in various disciplines (e.g., psychology, history, computer science, education). As a result, each IR practitioner is influenced by their formal training and their own experience when conducting studies. Conceivably, two IR practitioners conducting a study into a single area/program could yield somewhat different results based upon different approaches grounded in different disciplines. A theory of institutional research (or set of theories) can provide guidance in the selection of methods and how information is gathered and interpreted.

### **Purpose of the Study and Research Questions**

The purpose of this study is to map some Evaluation methods and approaches to IR. Both share similar beginnings, purpose, reactions, and concerns. The primary difference between the two professions is that evaluation is an established discipline. Thus, evaluation is afforded the forums to more deeply study its self. Consequently, approaches and models for evaluation have been developed, tested, challenged, and refined. As a result evaluations' self-

studies, IR can learn from evaluation's processes, procedures, methods, approaches, research activities, and discussions.

The primary objectives of the current study are summarized by the following research questions:

1. Does an Evaluation Theory Tree (Methods, Use, and Values branches)—the Evaluation Theory Tree is introduced in the next chapter—fit Institutional Research?
2. How do “Evaluators” differ from “Non-evaluators” in their placement on each branch of the Evaluation Theory Tree?
3. Which approaches to conducting IR are significantly related to Core IR reporting?
4. Is there a relationship between office FTE and branch scores?
5. Is there a relationship between institution type and branch scores?

## **CHAPTER II**

### **LITERATURE REVIEW**

#### **Overall Purpose**

The purpose of Evaluation is not universally agreed upon. Chelimsky (2006) provides three perspectives towards evaluation's purpose. The first is holding policy makers accountable for the merit and worth of their policies and programs. The second is generation of knowledge. Last is improvement to institutions. Patton (1996) adds at least seven purposes of evaluation that include: (a) generating knowledge and principles of program effectiveness, (b) developing programs and organizations, (c) focusing management efforts, (d) creating learning organizations, (e) empowering participants, (f) supporting and enhancing program interventions, and (g) critical reflection for enlightened practice. Still, many within the field of evaluation consider social betterment to be evaluation's ultimate purpose (Christie, 2007; Christie & Azzam, 2005; Henry, 2005; Henry & Mark, 2003; Lawrenz, Gullickson, & Toal, 2007; Mark, Henry, & Julnes, 1999; Mohan & Sullivan, 2006; Rossi, Freeman, & Lipsey, 1999; Sridhara, 2003; Stufflebeam & Shinkfield, 2007). Given government's role in requiring evaluation of its programs and services during the 1960's and since many of the influential evaluators of today are or have been involved in evaluation of government programs, the popularity of a social betterment

purpose seems somewhat predictive. Henry and Mark (2003) define social betterment as “the improvement of social conditions” (p. 295). Others include within this definition a judgment rendered through deliberation and public opinion (Henry, 2000; Henry & Julnes, 1998; Mark et al., 2000). Social betterment is intended to improve the lives of the underrepresented, less privileged, and disenfranchised. By bettering these targeted population’s lives, overall society improves too. Alkin and Christie provide a visual representation of the social betterment purpose of evaluation by depicting a tree whose trunk is built from a dual foundation of accountability and systematic social inquiry (Alkin, 2004). As defined in their context, accountability is social betterment: “accountability...is designed to improve and better programs and society” (p. 12). Thus, and according to Alkin and Christie, evaluation is built squarely upon the concept of social betterment.

Even with its apparent popularity, social betterment creates a limit on the scope of evaluation’s purpose. Consider that there are programs and organizations that do not have a social context. As a result, evaluation of such programs would not contain a social betterment outcome. Alkin and Christie help to clarify what is meant by social betterment by including “the enhancement of educational and social conditions through the improvement of programs and organizations designed to address these conditions” (p. 297) and that “evaluation should contribute to the decision-making process of organization members” (p. 345). Alkin’s “contribute to the decision-making process” statement is perhaps

recognizing the fact that evaluators do not make decisions to implement changes, but rather provide information to decision-makers who are responsible for such choices. This view, that evaluation provides information to decision-makers, is not uncommon among researchers (Cronbach, 1963; MacDonald, 1976; Mangano, 1991; Preskill & Caracelli, 1997; Stufflebeam et al., 1971; Weiss, Murphy-Graham, Petrosina, & Gandhi, 2008).

According to Patton (1996) summative and formative evaluation is about how findings are used and as a result limit the scope of evaluation's purpose. For example, going through the evaluation process may lead to a different mindset, a different way of thinking and not necessarily an improvement to the focus of the study. Scriven (1996) considers Patton's examples of evaluation as "consulting" and not evaluation. As an example, while a change in mindset or thinking is an important finding, this alteration appears to be an impact of the evaluation. Scriven (1996) considers teachers of evaluation as consultants—since they are not conducting an evaluation. Thus the evaluation can have an impact worth noting—particularly an impact upon the tangential. However, according to Scriven, the purpose of evaluation is to determine merit and worth. Scriven would probably agree that it is difficult to imagine that a purpose of evaluation is to simply create learning organizations or to focus management efforts, but rather it creates effective learning organizations and helps to focus effective management efforts through formative findings. Through these

“effective” efforts, evaluation might impact individuals to alter their thinking, mind set, or some other noble and note-worthy ancillary effect.

With such divergent views on the purpose of evaluation, is it possible to synthesis a single general purpose of evaluation that makes sense? When considering the previously expressed purposes and boiling those down to simplicity, it appears that evaluation is about providing information to decision-makers so that appropriate decisions can be made. In most cases the decisions would leave the area under study as is (i.e., the program is effective) or alter it (i.e., the program could be improved or terminated). This description lends itself to summative and formative evaluation (a judgment of worth/merit and modification, respectively). This description also lends itself to effectiveness. But, what exactly is effectiveness? Wordnet (2008) supplies definitions that might prove helpful in understanding what effectiveness means. For example, “quality of being able to bring about an effect” and “the quality of being effective” is provided. In addition, effective is defined as “adequate to accomplish a purpose; producing the intended or expected result” (Wordnet, 2008). The American Evaluation Association (2008a) provides verbiage that also raises an effectiveness purpose by stating, “Evaluation involves assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to *improve their effectiveness*” [emphasis added]. Based upon this evidence, one can argue that evaluation’s purpose is about effectiveness—that being it either proves or improves effectiveness.

While the purpose of IR appears to be less debated than evaluation, how IR is perceived is clearly debated in various articles. For instance, is IR considered a policy analyst (Gill & Saunders, 1992; Terenzini, 1993), an organizational intelligence officer (Tetlow, 1984; Terenzini, 1999), or adaptive function contributor (Peterson, 1999)? Regardless of how IR is perceived, there appears to be a somewhat consistent view on its purpose being that it provides information to decision-makers so appropriate choices can be made (Baker & Roberts, 1989; Delaney, 1997; Dressel, 1981; Ehrenberg, 2005; Fincher, 1985; Knight, Moore, & Coperthwaite, 1999; Peterson, 1999; Saupe, 1990; Teodorescu, 2006). Similar to evaluation, this description lends itself toward summative and formative findings and thus to an overall arching purpose of effectiveness. Once again and similar to evaluation there is the familiar theme of proving or improving effectiveness. Comparable to the AEA, the Association for Institutional Research (AIR), provides verbiage that supports this arching purpose of effectiveness by indicating that some of AIR's (2008) purposes are to

further the professional development and training of individuals engaged in institutional research and analysis or interested in its utilization in planning, management, and resource allocation and in *the improvement of postsecondary education* [emphasis added]

and to “advance research and analysis *leading to the improved* [emphasis added] understanding, planning, management, and operation of postsecondary educational institutions and agencies” (AIR, 2008). Based upon these findings it

appears that the purpose of IR is to prove or improve effectiveness—whether it is program, policy, personnel, product, or organization (institutional) effectiveness.

In summary, it appears that neither evaluation nor IR are in positions that make decisions to implement change, but rather provide information to decision-makers who are responsible for making such choices; each provides information that adds to a judgment of merit, worth, or modification to the area under study (summative and formative findings); and finally each has the purpose of proving or improving effectiveness at levels that vary from the specific (i.e. program or policy level) to the broad (i.e., organization or institutional level). Still, other professions have similar purposes and are very different in their methods and approaches in applying their craft. For example, accounting helps to run a business and so does management. Yet, it would be a stretch to say that these two professions are more similar than dissimilar. However, at least a first hurdle has been passed by establishing similar purposes between evaluation and IR.

### **History**

The study of history provides insight into the past and, if considered appropriately, can provide insight to the future. By doing so, greater clarity to differences and non-differences can surface that help to understand each profession better. For example, evidence can be gathered that might indicate when divergences or concurrences occurred and possibly answer questions or shed light on why paths took certain directions. Evidence of formal evaluation activities can be traced back hundreds and even thousands of years (Mertens,

2001; Rossi et al., 1999; Scriven, 1991; Shadish, Cook, & Leviton, 1991; Shadish & Luellen, 2005; Stufflebeam & Shinkfield, 2007). Shadish and Luellen (2005) indicate that evaluation can be traced by to biblical times through the Old Testament in the Book of Daniel via the comparative effects of a Hebrew diet versus a Babylonian diet on health. They provide additional, earlier evidence tracing evaluation activities to approximately 2200 BC in China and the nature of valuing for millennia. However, history varies depending upon perspective. When considering the perspective of the domain being studied such as, program, policy, personnel, product, or student evaluation, the history of evaluation becomes complex and more difficult to pin-point exact beginnings (Mark, Greene, & Shaw, 2006; Shadish & Luellen, 2005). Mark et al. (2006) cite examples of United States educational evaluation starting with Ralph Taylor during the 1930s, non-educational social program evaluation starting in the 1960s with the Johnson administration's Great Society, and Stafford Hood and others indicate that many histories of evaluation ignore contributions of African Americans—most likely due to discrimination. In addition, they indicate evaluation history has geographical bounds. Thus, evaluation history is contingent upon which country is being considered such as Israel, China, England, or the United States of America. Due to the complexity due to domain and country being studied, additional examples of the history of evaluation will be limited to that in the United States of America and of program evaluation. Program evaluation is identified by Shadish and Luellen (2005) as having the primary impact on the explosive growth in

evaluation theory and practice over the past 40 years. Program evaluation is defined as a formalized approach to studying the goals, processes, and impacts of projects, policies and programs.

Rossi et al. (1999) indicate that systematic evaluation is relatively new. Stufflebeam and Shinkfield (2007) add to this observation by indicating that systematic evaluation within a school system can be traced back to at least 1845 in Boston with oral examinations conducted by school committees. In 1845 Boston replaced oral exams with the first systematic school survey using printed tests. Horace Mann approached and convinced Boston officials to base their policies on test results from the eldest class in each of the city's nineteen schools. However, Joseph Rice is credited with conducting the first formal education program evaluation in New York City in 1895 (Stufflebeam & Shinkfield, 2007). Rice collected data over the next decade on spelling and math scores from approximately 16,000 students.

The 1930s seems to be a critical stage in evaluation's history. As a result of the Great Depression, President Roosevelt's New Deal created a number of relief agencies. Perhaps more than anything else, the New Deal created the precedence for federal support of social programs. During this same period, Ralph Tyler (often referred to as the father of educational evaluation) developed and refined an evaluation approach that was a clear alternative to other views (Stufflebeam & Shinkfield, 2007). His approach concentrated on clearly stated objectives and measured whether or not those objectives were achieved. This

had a clear impact on curriculum and test development. To Tyler's benefit, he and Smith were commissioned to conduct the Eight-Year Study (Smith & Tyler, 1942). This research conducted on 30 schools across the nation allowed Tyler to expand, test, and demonstrate that his approach to educational evaluation had clear advantages over others. Tyler's approach was so accepted that it influenced how program evaluation would be viewed for the next 25 years.

While the 1930s can be considered a critical stage in evaluation's history, the 1960s tends to be recognized as its professional beginnings (Mertens, 2001; Rossi et al., 1999; Shadish et al., 1991; Shadish & Luellen, 2005; Stufflebeam & Shinkfield, 2007). Following President Kennedy's assassination, President Johnson followed through Kennedy's expanded social programs in his "Great Society" plan. However, budgets are limited and there were competing interests for funding. During the Great Society time period, the United States was increasing its commitment in the Vietnam War, education in math and sciences were expanded to counter the effects of the Soviet Union's launching of Sputnik, the Cold War is well under way, inflation was increasing, oil embargoes were occurring, NASA was trying to put a man on the moon, and federal budget deficits were increasing. While many of these initiatives were geared towards national security, the expanding social programs were under scrutiny to prove that they are achieving expected outcomes. Evaluation was one mechanism for that proof. As such, evaluation became attached and mandated into federal grants. Thus, to receive federal grants an evaluation component became

required. As a result, evaluations' professional beginnings can be pegged to the mandated requirement from the federal government.

Evidence of formal IR activities can be traced back a few hundred of years. IR functions are typically embedded within educational institutions. This creates a limiting factor to its history. For instance, in the United States of America, this would limit the time frame to perhaps the oldest operating established institution—possibly Harvard (established in 1636). Outside the bounds of the United States of American, the time frame would be limited to perhaps the University of Bologna in Italy established nearly a thousand years ago in 1088 and still operating today. However, Cowley (1959) cites the origins of IR to self-studies in the early 1700s and the founding of Yale. While the evidence on Yale appears to be the earliest documented activities of IR in education, Cowley indicates that collections of quantitative data first emerged in the 1870s at Harvard.

Mortimer and Leslie (1970) indicate that something approximating IR occurred between 1910 and 1920 when bureaus of educational research were established in public school districts and universities (as cited in Dressler, 1981). Interesting enough, many of the early studies conducted were concerned with deficiencies in writing, reading, and math. These deficiencies mirrored Rice's concerns in his study in New York City in the late 1800s to early 1900s. Generally speaking, one major difference between educational evaluation and IR appears to be that IR studies are conducted on college students and educational

evaluation studies are based on K-12 students. IR and evaluation overlapped during this time—both conducting studies addressing the same concerns and deficiencies, but on different populations.

In the 1920s to 1930s, self-studies as part of accreditation efforts lent itself toward IR. Although self-studies are considered a minor role for IR (occurring once every ten years by regional accrediting agencies), effective and efficient self-studies requires continuing collection and analysis of data during the interims between formal self-studies (Dressel, 1981). As accreditation evolved into specific program areas (such as business, communication, chemistry or music) IRs role with respect to accreditation also increased to help acquire specialized program accreditation status. Although the regional accreditation status is necessary for federal funds, specialized accreditation is less of an issue and more dependent upon individual school needs and wants.

The overlap between evaluation and IR can be illustrated through the work of Ralph Tyler in the 1930s to 1940s. His work blends nicely into IR. At least some of Tyler's work is recognized as IR by Dressel (1981). This recognition should come as no surprise given Tyler's designation as the father of educational evaluation. Again, the difference in recognizing studies as evaluation versus IR appears to be based upon the population being studied (college level students versus primary and secondary level students).

Many within IR point to growth and national meetings involving IR as a topic during the 1960's as important and possible formative years (Dressel, 1981;

Fincher, 1985; Saupe, 2005; Tetlow, 1979). Doi (1979) provides a contextual growth pattern by providing a time line for the expansion of IR. Doi charts the American Council on Education's creation of the Office of Statistical Information and Research and the reports that flowed from it, the impact of the GI Bill of Rights in 1944, the 1959 Western Interstate Commission for Higher Education-Stanford workshop on college self-study, and the Southern Regional Education Board three year series of institutional research workshops that began in 1960 as catalysts for IR. However and less recognized is the Higher Education General Information Survey (HEGIS) impact on IR. Implemented in 1966, these federally requested reports should also be considered a viable starting point for IR as well. During its early years and much like today, IR is considered an institutional reporting function—supplying aggregate information to both internal and external audiences (Fincher, 1985; Lindquist, 1999; Muffo, 1999; Peterson, 1999; Volkwein, 1999). HEGIS was the first systematic federal reporting of all recognized accredited institutions. Institutions reported information via a number of standard surveys. These surveys collected institution-level data on such topics as institutional characteristics, enrollment, degrees conferred, salaries, employees, financial statistics, libraries, and others (National Center for Education Statistics, 2011). While HEGIS reports were not mandated, institutional compliance were tracked and delinquent schools were contacted via Western Union telegrams and requested to provide an estimated completion date within one week of receiving the telegram. In 1987 HEGIS migrated into the

Integrated Postsecondary Education System (IPEDS). Eventually the federal government mandated submission of IPEDS reports in 1992 and would fine delinquent institutions (Brown, 2008).

When evaluation moves into the education arena, the lines of distinction between IR and evaluation become blurred. Although, the term “evaluation” appears to be used for the study of K-12 grades and the term “institutional research” appears to be used for the study of college. Again, the terminology use appears to be more of a function of the population being studied than from true differences between the two professions. How terminology based upon the population being studied came to be is unclear at this point. The federal government holds a major role in the flourishing of each profession. During the 1960s, federal mandates for evaluation to be included into government grants and the highly requested college HEGIS surveys for IR are considered pivotal requisites in developing, promoting, and growing each profession. When dating systematic evaluation and formalized IR operations to approximately 1900, both professions are considered relatively new. However, if the 1960s are considered the “flourishing years” for evaluation and IR, both professions are much newer than the scant systematic/formalized period of the early 1900s. Additionally, it might be of importance that at least some of Ralph Tyler’s work is recognized as evaluation and IR—particularly since he is considered the father of educational evaluation. Although it is difficult to say that history provides clear indications of a divergence or concurrence of the professions, it does provide interesting and

possibly compelling reasons to explore the similarities between evaluation and IR.

### **Educational Backgrounds**

Given the similar and overlapping history of both professions, resemblance in educational backgrounds may not be that surprising. The educational background information presented here was collected through surveys from individual researchers and not as a function of a standard, on-going collection effort via professional membership organizations. A 1993 survey of 2,045 America Evaluation Association (AEA) indicates that their membership's education is the following: Education (22%), Psychology (18%), Evaluation (14%), Statistical methods (10%), Sociology (6%), Economics and political science (6%), Organizational development (3%), and Other (21%). This clearly shows the diverse educational background of AEA members (Rossi et al., 1999).

Lindquist's study (1999) examined a number of surveys conducted on Association for Institutional Research (AIR) in addition to the 1998 AIR membership database to obtain attributes of members. Findings from that study indicate that AIR members also have diverse educational backgrounds: education (about 40%), social sciences (about 30%), physical sciences (10% to 15%) and business (10% to 15%). However, some of the AEA categories can be combined into the comparable AIR categories so that a more accurate comparison can be made between the two membership organizations. For example, many evaluation programs are located within the education division,

and the social sciences tend to incorporate sociology, economics, history, and human services. A regrouping of the AEA results yields the following: education (36%—includes evaluation), social sciences (30%—includes economics, psychology, sociology, and political science), physical sciences (10%—includes statistical methods), and business (24%—includes organizational development and other). These new AEA figures mirror those of the AIR findings.

Both professions have members with diverse educational backgrounds. One question that cannot be answered from the Rossi and Lindquist studies is whether there is an overlap in membership (dual membership)? Answering this question might prove useful in comparing both professions more deeply. If a significant number of researchers hold membership in AEA and AIR, this might provide additional insight into the connections between evaluation and IR. For example, dual membership might provide evidence of the recognized benefits of being a member to both organizations or it might indicate the perceived overlap between the two professions.

### **Contextually Driven**

Evaluation is contextually based upon the program being studied (Bickman, 1994; Datta, 2007; Fetterman, 2001; Fitzpatrick, 2004; Greene, 2005b; Greene et al., 2007; Julnes & Rog, 2007; Scriven, 1996; Thomas, 2000). The Joint Committee on Standards for Educational Evaluation (JCSEE) explains the influence derived from contextual settings, “The context in which the program exists should be examined in enough detail, so that its likely influences on the

program can be identified” (1994, p. 125). Fitzpatrick, Christie, & Mark (2009) reaffirm the importance of context by indicating that evaluation practice is situational. Thus a sufficient awareness and understanding of the contextual setting should be understood by the evaluation team so that its impact can be incorporated into the study. Greene (2005b) indicates the comprehensive importance of context by indicating that “*all* [emphasis added] evaluation theories and the challenges of context are inescapably present in *all* [emphasis added] evaluation practice” (p. 82). According to Greene (2005b), context refers to the “setting within which the evaluand (the program, policy, or product being evaluated) and thus the evaluation are situational” (p. 83). Mark (2001) adds context’s description by adding the following: the features other than the intervention, the attributes of clients, the setting of the evaluation, the historical moment, and the service deliverers. In addition, Mark, citing Pawson and Tilley (2001), indicates the importance of episodic events whose significance needs to be considered as well. As a result, all evaluations should be respectful of, responsive to, and tailored to its contexts (Greene, 2005b, p. 84).

The contextual setting can be a complex array of attributes. Fitzpatrick (2004) provides examples that indicate context affects the role of the evaluator (internal versus external), the breadth and depth of stakeholder involvement, the methodology selected, and even the use of results. Contextual attributes for further consideration are the political settings, utilization of findings, organizational structure, and even social conditions of the evaluation. Fitzpatrick

(2009) continues to indicate the importance of context by positing that successful evaluators are aware of the need to consider the characteristics of the context. Chen (1996) and Grob (2003) also put forward that examination of contextual factors allows for the selection of the type of evaluation that is best and that most useful evaluations include contextual settings that allow for the reader to “put things in perspective.”

While context’s importance is recognized, it is also acknowledged as a limitation of any evaluation study—i.e., each evaluation is or should be tailored to the situation at hand. This leads to questions about context within the frame work of generalizability. One solution is the use of meta-analysis. While evaluators employ this method, there appears to be an adequate amount of researchers who do not record sufficient contextual information and thus, hinder greater use of meta-analysis (Lipsey, 2001). Mark (2001) addresses the generalizability issue in evaluation by proposing to make it “an empirical question”—letting patterns of evaluation findings provide answers rather than assuming the answers in advance. While its importance is often cited, contextual settings are complex and it is this complexity that leads to subjective inclusion of information in evaluation studies. At some level, this subjectivity creates questions on how to best account for contexts’ impact.

IR duties, responsibilities, and studies are shaped by the institution where it is housed and thus the institution being studied (Chase, 1979; Chase & Tetlow, 1979; Fincher, 1985; Lane & Brown, 2004; Peterson, 1985, 1999; Schmidlein,

1999). Thus IR is also contextually based. Lane and Brown (2004) indicate that context can have a large impact on institutional performance and decision making. As a result, institutional context has an inherent connection to understanding measures that address accountability within higher education. For example, some performance measures use graduation rates as an indicator. It is commonly known that the more academically prepared students are when entering college the more likely they are to graduate. It is also commonly known that highly selective colleges tend to have higher graduation rates than less selective colleges. Yet, knowing that College A is a highly selective institution and College B is a less selective institution provides greater meaning than simply stating that the six year graduation rates for College A is 90% and College B is 45%. Thus, context adds greater understanding to numbers.

Knowing and understanding various college types can add to the complexities of contextual settings. Service academies and tribal colleges provide examples of the importance to understanding the contextual differences among institutions. At service academies, graduating students are required to enter the service for five years. In essence, these students know what they will be doing for the next nine years of their lives—four years at the service academy and five years of military service. In addition, each student is expected to excel in three general areas: academics, military training, and physical education (Forest, 2003). These students also tend to be well prepared academically and are dedicated to the principles of service and discipline (Forest, 2003). Tribal

colleges present an array of contextual settings. These colleges range from technical schools to baccalaureate degree granting institutions. As a result, student educational goals range from employment training to entering graduate school (Ortiz & Boyer, 2003). In addition, many of these schools are remotely located with direct linkages to the K-12 local system. As a result, understanding the K-12 system for these colleges becomes a necessary attribute to better understand the challenges associated with tribal colleges (Ortiz & Boyer, 2003). There are many college types and thus, many contextual settings, that need to be considered when studying higher education and all its complexities. Some of the varying college types are provided with an accompanying resource for further understanding contextual settings: community college (Lane, 2003), historically black college or university (Brown, 2003), women's college (Langdon & Giovengo, 2003), virtual classroom (Kinser, 2003), corporate schools (Allen & McGee, 2004), religious affiliated college (Smith & Jackson, 2004), proprietary schools (Zamani-Gallaher, 2004), professional school (Sun, 2004), theology schools, or transnational campuses (Lane, Brown, & Pearcey, 2004). Comparing colleges simply by the "numbers" with no contextual understanding of each type of school can lead to results that are misleading.

Even with the current research on context, Lane and Brown (2004) indicate that IR is lacking research on it within its literature. Perhaps this due to the naturalistic setting in which IR resides—that being IR is most often conducted within a single institution (Dressel, 1981; Fincher, 1981, 1985; Lane & Brown,

2004; Peterson, 1999; Saupe, 1990; Schmidlein, 1999) and typically on a particular aspect of that institution such as grade inflation or retention. As a result, the contextual influences in IR become ingrained, are often unaccounted for, over looked, or simply understood as common knowledge. Volkwein (1999) provides descriptions of what he views as the four faces (roles) of IR (information authority, policy analyst, spin doctor and scholar and researcher). Within each face are descriptions that clearly relate to contextual influences upon IR. For example, in the information analyst role the institutional researcher describes the institution in its various descriptive forms (students, faculty, staff, buildings, financials, and activities) and that IR acts to educate the campus community about itself. It is clear that IR is exercising its information analyst position to assist and inform internal and (most likely) external stakeholders about the *institution*. In the policy analyst role the institutional researcher is educating senior staff/cabinet members providing support for planning, budgeting, institutional policies, administrative structures, and other institutional needs. In the role of spin doctor the institutional researcher is providing information that reflects favorably upon the institution. And finally in the role of scholar and researcher the institutional researcher investigates and produces evidence so that effectiveness, legal compliance, and goal attainment can be judged on the institution. Support for the institutional self-study is also cited. Although each face is inherently connected to contextual influence—IR is conducting studies, producing reports, and supporting decision making at their particular institution—

Volkwein never uses the word “context” in this article. This is likely due to the general nature and applied working conditions of IR.

Terenzini (1999) indicates “context,” but does so in the form of organizational intelligence (OI). He describes the professional characteristics of effective institutional researchers and places these descriptions into one of three tiers and in doing so he presents evidence of the contextual nature of IR. Terenzini conceives IR as institutional intelligence in three distinct, but mutually dependent tiers. Tier 1 is technical and analytical intelligence and is considered the “basic building blocks” for IR. Tier 2 is issues intelligence. Tier 2 involves most of the substantive problems that tier 1 skills are employed. Tier 2 comprises knowledge of the major issues that the institution faces. IR is sensitive to and understands the major categories of problems confronting the institution’s administrators. This tier is directly related to supporting decision making at the institution. Commonly associated activities with Tier 2 are assessment, program evaluation, faculty evaluation, and the institutional self-study. At this tier, institutional researchers understand how the institution functions and how decisions are made. Tier 3 is contextual intelligence and involves the understanding of the culture of higher education in general and of the specific campus of the institutional researcher. Terenzini calls this the “crowning form of OI.” This tier reflects savvy and wisdom, and where IR has legitimacy, trust, and respect. This tier is seen as the pentacle in OI. It is obvious from Terenzini’s

descriptions that two of the three OI tiers have a contextual knowledge needed for successful IR.

Regardless of the reasons why IR literature lacks studies on the impact of context (i.e., the inherent nature of IR or embracing it for further study), its current research findings appear to address context from a different perspective than that of evaluation. The IR literature appears to focus upon the mission of the program under study. For example, IR research considers the mission and, thus the type of institution (i.e., tribal colleges, community college, or single sex colleges). Evaluation appears to include attributes surrounding the program of study (i.e., method proclivity, stakeholder involvement, internal-external role, and summative-formative evaluation focus). In comparing the two professions, the nature of IR can easily account for the lack of studies within the framework of the internal-external roles and summative-formative types of studies. For example, most institutional researchers would not be considered an external evaluator. IR is more aligned with the departmental evaluator (internal to the organization but external to the program understudy) as described earlier by Cummings et al. (1988). As a result, researching the differences among roles for IR is somewhat mute because there is only one role to study - the departmental role. In addition, a summative rendering is not normally within the purview of IR. Typically, IR offices evaluate programs and services from an improvement position (formative) and not from a summative position (judging the merit or worth). Finally, higher education tends to be a highly collaborative environment. As a result, stakeholder

involvement is normally, although not always, a given. As a result, this type of context measurement could be overlooked in IR studies because it is assumed or implied that stakeholders are involved. Still, there are varying levels of stakeholder involvement that could be accounted for in IR studies (i.e., none, little, moderate, quite a bit).

Lane and Brown (2004) consider context from three main components to help aid IR in conducting studies: artifacts, implicit assumptions and location. Artifacts are considered the most visible and are derived from the physical and social environment (Schein, 1985). Artifacts convey information about and culture of the organization. Aspects of artifacts are language, values, beliefs, ways of doing, ceremonies, and traditions. Implicit assumptions tell groups how to perceive, think, and feel about things. So implicit are the assumptions that many within the organization do not realize why they engage in certain things the way they do. Finally, organizations are influenced by the culture of the community in which they reside. Thus location influences the organization.

When considering the research conducted on the influence of context within IR and Evaluation, the evidence indicates that it is important. However, Evaluation and IR have different perspectives when considering how context is viewed. Evaluation tends to focus on attributes that surround the program of study (such as level of stakeholder involvement or methodology selection) and IR tends to focus on more general concepts (such as mission of the program or

organizational type). It should be clear that both professions recognize that each is contextually driven.

### **Method Proclivity**

Contextual setting recognition has led to the development and improvement of methods, procedures, and approaches that enhance identification and acknowledgment of its importance in designing and conducting an evaluation. One of the most evolved and popular evaluation approaches is program theory-driven evaluation (Donaldson, 2007). Donaldson (2007) indicates that a major contribution to program theory-driven evaluation is its emphasis on working with relevant stakeholders to tailor the evaluation to meet agreed-upon values and goals. This popular evaluation approach is highly sensitive to contextual settings. As presented by Greene (2005a), the value-engaged approach devotes a great deal of time to contextual learning or the front-end aspects of evaluation. Greene includes the following transferable attributes from her study of the Bunche-Da Vinci Learning Center evaluation; learning the context and the program to be evaluated, developing appropriate relationships with key stakeholders, understanding the critical issues to be addressed, identifying priority evaluation questions, and determining criteria for making judgments of program quality. As a result, the evaluation blends into the context. Greene indicates that only after the contextual attributes are absorbed can an evaluation design be designed and implemented. Many, if not all, evaluation approaches focus some level of attention towards contextual understanding—

such as Empowerment Approach (Fetterman, 2001), Results-Oriented Management (Wholey, 2001), and Utilization-Focus (Patton, 1997).

Even though individual evaluation researchers are drawn to particular methods, research findings indicate that contextual settings often affect the methodology employed (Fitzpatrick, 2004; Greene, Lipsey, Schwandt, Smith, & Tharp, 2007). Greene et al. (2007) indicate that decisions about methodology occur due to an intersection with theory and the particular contexts. Although Fitzpatrick (2004) indicates that exemplar studies appear to use a mix of qualitative and quantitative methods, her observation is qualified by recognizing that all of her examples are multiyear studies—an uncommon situation for most evaluations. Thus the duration of the evaluation may have an impact on the depth and breadth of methods used. Contextual settings impact has lead to further discussions about methodology. For example, some federal programs appear to promote experimental designs (in particular randomized studies) in research and evaluation (Julnes & Rog, 2007). While randomized experiments have been referred to as the “gold standard” (Datta, 2007; Greene et al., 2007), promoting such a methodology could inherently disregard contextual settings; thus, a more appropriate methodology is not considered. In addition, randomized experiments have ethical considerations to consider; such as placing human subjects deliberately into groups were treatments are expected to not return improvements. As a result, there is no one best method to be applied to all

evaluations. Perhaps Trochim's (2004) statement best summarizes how to answer which method is best— "It all depends."

McLaughlin, McLaughlin, and Muffo (2001) most directly address contextual influence upon method proclivity/selection within IR by indicating that different groups need information generated through the use of different methods. They further explain methodological selection through a case study where "concerns of the audience needing the information also had to be factored into the choice of methods" and that "clearly, no one analysis or study or methodology will be seen as persuasive by all of these groups" (p. 20). Thus the contextual setting impacts method selection. While research into method proclivity within its literature is limited, IR appears to address contextual influence from a functional perspective. An important function of IR is to provide information in such a way so as to reduce uncertainty for decision makers (Borland, 2001; Howard & Borland, 2001; Saupe, 1990). Reducing uncertainty can occur by employing multiple methods, a triangulation approach, where different methods yield confirming results. Confirming results from multiple methods provides additional information so that decision makers can come to closure. Effective IR considers context when selecting methods so that its studies have value to decision makers. To a certain extent, Terenzini (1999) addresses contextual influence and method proclivity in his third tier of intelligence (contextual intelligence) by indicating that IR must know how to effect change. Understanding how to effect change involves understanding which

methodologies are acceptable, understandable, and needed to help reduce risk for various decision makers. Reducing uncertainty means understanding that the contextual setting effects method proclivity.

### **Stakeholder Involvement**

Context also influences depth and breadth of stakeholder involvement. Fitzpatrick (2004) cites examples where the scope and context influences the nature of stakeholder involvement. If the setting allows for more or less stakeholder input, the evaluator adjusts accordingly. A 12-day study doesn't allow for the depth and breadth of stakeholder involvement as does a 12-month study (should the primary stakeholders allow or expect for such involvement). At other times the political setting allows for limited stakeholder participation. However and perhaps most important is that the exemplar studies presented by Fitzpatrick (2004) all exhibit a central theme of stakeholder involvement. Still the stakeholder involvement does bring concerns. Some researchers advocate stakeholders in the design and conduct of the evaluation (Green, Mulvey, Fisher, & Woratschek, 1996; Knapp, 1995). While this level of stakeholder involvement is intended to help reflect diversity, tailor suggestions for improvements, or identify and assess program outcomes, there is evidence that stakeholders lack assessment skills such as construct definition and instrument development (Impara, Plake, & Fager, 1993). Deficiencies in skills range from outdated to no knowledge, or a lack of experience in this area of study. Determining the purpose of stakeholder participation can be critical if not a necessary step in the

evaluation design. Non-collaborative evaluations with respect to stakeholder participation are typically conducted for the purposes of generating validity of findings (Brandon, 1998). While previous research findings indicate that extensive (collaborative) stakeholder participation tends to be conducted with the primary purpose of enhancing use of findings (Bryk, 1983; Cousins, Donohue, & Bloom, 1996; Cousins & Earl, 1992; Johnson, Willeke, & Steiner, 1998), Brandon (1998) presents evidence of collaborative evaluations that can increase validity. Still, the distinction between collaborative and non-collaborative stakeholder participation is not always clear (Brandon, Lindberg, & Wang, 1993; Brandon, Newton, & Harman, 1993). A non-collaborative evaluation could utilize stakeholders at the early stages of the evaluation to better educate the evaluator. During this period, stakeholder involvement can become very intense and can have a long duration; thus blurring the difference between a collaborative and non-collaborative evaluation.

Brandon (1998) provides examples of the win-win nature that some collaborative evaluations generate—that being enhancing the likelihood to produce valid findings and enhancing the use of the evaluation results. Brandon followed Smith's (1997) suggestions and employed three broad procedural rules in collaborative evaluations that enhanced validity (as cited in Brandon, 1998). Rule one, tapping the expertise of the appropriate group should be obvious that this is a way for the evaluator to gain knowledge about the history, purpose, processes, and procedures of the program. These stakeholders have first-hand

knowledge about program features that are working well, need alterations, or need to be terminated. These stakeholders are represented by program staff, managers, and beneficiaries. Rule two, tapping the expertise fully and carefully has more to do with the validity of applying carefully developed and thorough methods for stakeholder participation. Brandon describes a very thorough process that employed a Delphi-like procedure to develop a questionnaire, administration of the questionnaire, and then holding meetings with stakeholders to reconcile group differences. Brandon goes on to describe a face-to-face decision making process for furthering enhanced validity. This process brings groups together in the same meeting to help reconcile differences. Most important in this process is providing clear and easily understood group results, equal time to all groups to voice concerns/questions/opinions, and summarizing stakeholder comments throughout the meeting while ascertaining that its content is understood by all participants. Brandon provides a possible format for groups whose input had been routinely ignored by involving stakeholders in meetings to review and possibly revise evaluation recommendations. The structural format for such meetings tend to be brief and convenient, and recommendations are jargon free, carefully edited, and presented as chart essays. The third rule is involving stakeholders equitably—both between and within groups. Different stakeholders can influence evaluation decisions differently based upon job position and personal characteristics (Bacharach & Lawler, 1980). The primary intent of rule

three is to ensure that no one stakeholder groups' expertise is ignored in evaluation decision making.

One possible method to ensure that no particular group or person dominates the meeting is to employ the Nominal Group Technique (NGT) developed by Delbecq and VandeVen (1971). Like any technique, NGT has its advantages and disadvantages. This technique tends to encourage more passive group members to participate by providing anonymous answers/suggestions and equal participation time. Still, some might find this process too mechanical and ideas may not converge. While more extensive readings on this technique are encouraged, the following steps are sufficient to understand the process (Sample, 1984):

1. Divide the people present into small groups of 5 or 6 members, preferably seated around a table.
2. State an open-ended question ("What are some ways we could encourage participants to car pool?").
3. Have each Person spend several minutes in silence individually brainstorming all the possible ideas and jot these ideas down.
4. Have the groups collect the ideas by sharing them round robin fashion (one response per person each time), while all are recorded in key term, on a flipchart. No criticism is allowed, but clarification in response to questions is encouraged.

5. Have each person evaluate the ideas and individually and anonymously vote for the best ones (for example, the best idea gets 5 Points, next best 4 Points, etc).
6. Share votes within the group and tabulate. A group report is prepared, showing the ideas receiving the most points.
7. Allow time for brief group presentations on their solutions.

Just as method proclivity is viewed differently by IR, so too is stakeholder involvement. IR tends to associate stakeholder involvement as nearly a given due to the nature of higher education. Consider that higher education has a relatively long history of being governed by a dualism of control structure (Birnbaum, 1988) where faculty members, along with college administrators, participate in the governance of an institution. While the level and degree of participation varies between the two groups by school, higher education within the United States has operated quite effectively and efficiently under this structure for quite some time. Understanding why dualism of control exists in higher education may help to understand the complex and at times perplexing power structure in this industry. Dualism exists primarily from two reasons: (a) how higher education views itself, and (b) faculty governance. Higher education views itself as a community of colleagues or a collegium (Birnbaum, 1988). Collegiums have members that are viewed and accepted as equals. As a result, all members have equal input, voice, and valued knowledge. The American Association of University Professors (AAUP, 2006) Statement on Government of

Colleges and Universities document clear outlines a dualism of control structure that specifically outlines faculty role in institutional governance. AAUP plainly cites that joint effort (dualism) increases the possibility to solve educational and institutional problems. Thus, higher education typically operates under the collegial group decision making process (Borden & Delaney, 1989). As a result, it is important to IR professionals to understand group decision making processes, styles, and needs (Ewell, 1989). Successful groups tend to have (a) informed members, (b) members with differing perspectives, and (c) authority to arrive at solutions (Borden & Delaney, 1989). It should be clear that IR operates within collegiums.

### **Internal-External Role**

Another element of context is the role of the evaluator. Is the evaluator internal or external to the program or organization under study? Before categorizing evaluators, it may be helpful to understand the difference between external and internal evaluation. Ray defines external evaluator as someone who is not on the staff of the organization being evaluated and is contracted for employment for a specific period of time. As defined by Kendall-Tackett, an internal evaluator is “any staff person directly involved in the program under evaluation, or in the agency in which the program is housed” (as cited in Yang & Shen, 2006). By these definitions, it might seem obvious as to what constitutes an external evaluator. However, at other times it is not so clear. Determining the extent of the relationship and thus the role of the evaluator is more complex and

less agreed upon than one may think. Ray provides evidence of this from the 2002 annual AEA conference. A panels' response to the question how long an external evaluator can contract with a single agency and still be considered external and independent ranged from two to five years. The panel further explained that after some period of time the evaluator can become fiscally dependent upon the contract and so integrated into the staff that he or she is no longer really independent. As a result, an external evaluator who has too long a relationship and who has become financially dependent upon the contract can shift from being "technically" an external evaluator to that of a quasi-internal evaluator and thus an internal evaluator.

The advantages and disadvantages to both are well documented and well argued. For example, some of the more commonly cited advantages for external evaluations are objectivity and freedom from pressures of the organization (Ray, 2006), Worthen & Sanders add an impartial and fresh perspective (as cited in Yang & Shen, 2006), increased evaluation expertise, honest responses from participants (Russ-Eft & Preskill, 2001), and greater credibility of findings (Sonnichsen, 1987; Torres, Preskill & Piontek, 1997). Advantages assigned to internal evaluators are: understanding the contextual setting of the evaluand, greater access to data, tailoring the evaluation to the organization, sustainability of evaluation, reduced costs, greater probability of use (Russ-Eft & Preskill, 2001), and retention of knowledge (Patton, 1986).

Although most of the literature written about the role of evaluators has been from the perspective of the external evaluator (House, 1986), there appears to be an evolutionary shift from external to internal evaluations. In fact, House (1986) calls this shift “huge.” Torres, Preskill, and Piontek (1997) findings add confirmation to the shift by indicating that 52% of respondents to a randomly sampled survey of AEA members are internal evaluators. House (1986) contributes this shift as a result of evaluation being recognized as a “valuable commodity in society” and that administrators need to justify their decisions versus making decisions. Love’s (1983, 1991) findings tend to indicate that the growth in internal evaluation is a function of government mandating more evaluation and increases in available evaluation training. Regardless of the reason(s) for the shift, internal and external evaluators have different relationships and possibly different influences with the organization they are employed. House (1986) boils down the typical external evaluation process to an agreement being reached between the evaluator(s) and the sponsor(s) of the evaluation, the evaluation is conducted within accordance of the agreement, and because of an arms-length relationship between evaluator and sponsor, the evaluator produces an unbiased report. House’s concern is when the relationship changes from external to internal, the process can change. Internal evaluator’s careers can hinge on the repercussions of the evaluations. House (1986) outlines a number of concerns with internal evaluators. For examples, withholding of relevant data and findings or an occasional scandal can occur. In addition and

due the daily contact with administrators, the internal evaluator could be seen as a tool of the administration. If others perceive this “tool” relationship, the evaluator will be seen as doing what is in the interest of the administration and not that of the organization (House, 1986) and lose his or her credibility. The routine and repetitive nature of internal evaluation procedures can lead to a standardization of reporting. As a result of this standardization, reports could lose their meaning. In addition, internal evaluators could become “data-wise” (House, 1986) and circumvent vulnerable issues by cleverly answering questions or avoiding sensitive issues. Patton (1987) explains that a disadvantage for one can also be a disadvantage for the other. For example, internal evaluators can face pressure to produce favorable findings and external evaluators can also face the same pressures for positive results so they can obtain future evaluation contracts. Thus, one type of evaluator is not immune from the influences that the other faces.

Patton’s (2008) interview with ten internal evaluators provides empirical and compelling evidence of the realities that internal evaluators face in their work environments. Five general themes emerge from those findings that might prove useful for current or future internal evaluators. The first theme is that stakeholders might not engage in the evaluation because evaluation is perceived as the job of the evaluator. Thus it is the job of the evaluator to do the evaluation and not facilitate it. Second, internal evaluators are asked for public relations information rather than evaluation. Third, internal evaluators are asked for lots of

data-gathering and report-writing tasks that very time consuming, but too minor to be considered meaningful evaluation. Fourth, internal evaluators are often removed from major decisions or are too far removed from critical information networks to be appropriately used in the process. Fifth, getting evaluation results used requires lots of follow-through. Fortunately, Patton also provides examples on to overcome some of these issues. For example, finding and providing incentives for increasing stakeholder involvement and finding key internal and external persons to comprise an evaluation committee to help keep evaluation meaningful and to help derail “public relations” evaluation.

Cummings et al. (1988) findings from interviews conducted on three practicing evaluators provide interesting comparisons. Each evaluator interviewed represents each possible role of an evaluator; internal to the program of study, internal to the organization—yet independent of the program (referred to as departmental), and external to the organization. All three agree that if evaluation is not a sporadic function, then internal evaluation can be a cost savings. Each agrees that internal evaluation is not necessarily subjected to methodological limitations. All three agree that influences can be different between the internal and external roles; however, there is also agreement that influences can be transferable across roles. While routinization of reports can be cited as a possible concern (i.e., missing important data/information due to a failure to implement a better approach), all three recognize that routinization is not necessarily pejorative (such as longitudinal analysis, access to knowledge,

and reduced project time). In addition, all three indicate that quality evaluations are a function of evaluator expertise. The appearance of objectivity resulted in greatest differences. While it appears that objectivity can be built over a long period of time for internal evaluators, creditability is reduced by virtue of position when the clients are users who consume the product or service.

Still, there are benefits from an internal evaluator. An internal evaluator who understands the program, the culture, the politics, the organization, and other contextual settings can be a strong advocate to convert findings into action (House, 1986; Sonnichsen, 1987). Sonnichsen (1987) describes internal evaluators, with proper support as change agents practicing advocacy evaluation. Advocacy evaluation can be viewed as an ethos of evaluation that permeates the organization. This implies a belief and support of the value from evaluation. While Sonnichsen appreciates the potential concerns cited by House, as a practicing internal evaluator, he provides evidence from his own experience of the quality of internal evaluators and influential impact those findings have on organizational direction and policy.

In nearly all circumstances, the IR office would fit the departmental roles as described by Cummings et al. (1988)—internal to the organization, but external to the area under study. While many publications address the nature of IR as being a departmental role, we have yet to find publications that directly address this topic. Still, there are many publications that describe the role of IR such that one can easily arrive at Cummings departmental description. Saupe

(1990) indicates that "...institutional research may, intentionally or incidentally, identify situations within the institution which are causes for concern." This implies that while IR is internal to the institution it can also be external to the area under study. Lindquist (1999) provides examples from the two Primers on Institutional Research (1987, 1992) and AIR membership survey results (1981). Topics from the 1992 Primer included student persistence, enrollment management, student impacts, faculty demand, faculty salaries, peer institutions, diversity, environmental scanning, Total Quality Management, academic program review, and cost analysis. It is important to note that each topic area can be drilled down from the institutional level to other echelons that can include details at the school, division, or individual department/unit level. Still, some IR offices are located within government agencies or system offices (Lindquist, 1999).

### **Summative-Formative**

Scriven (1996) provides another way to view context influence by indicating that it also attributes to the summative and formative purposes of evaluation. Scriven (1996) explains how contextual settings alter summative and formative evaluation. He affirms that the distinction between formative and summative evaluation is a matter of roles and that roles are defined by context. Thus, context affects the type of purpose for the evaluation. He provides an example of customer's evaluation of a restaurant's soup. Summative results of the soup evaluation could indicate to keep the soup as is or that it is "hopeless." However, if the summative results are used to alter the soup recipe then the

evaluation takes on a formative function. Changing the context of the evaluation can change the evaluation from summative to formative and vice-a-versa. As a result, context has a major role in helping evaluation achieve its purpose. In addition, Scriven reaffirms that it is important to separate evaluation from action—that evaluation is rarely the only input in decisions. The importance of the contextual setting of the evaluation cannot be over stated—the contextual setting drives numerous, if not all, aspects of the evaluation processes and procedures. Context affects the methodology and approach, stakeholder involvement, purpose, use of results, role of the evaluator, and is a catalyst for developing and improving evaluation methods, procedures and approaches.

Volkwein (1999) addresses concerns of summative approaches in higher education (and thus IR). He specifically addresses problems associated with summative findings that may not address the internal development and educational enhancement—thus a formative approach to IR. Volkwein further indicates that higher education has a

renewed interest in process measures, rooted in the theory that good outcomes will not result from flawed educational processes. Measurement at critical process points enables institutions to determine which student experiences are having the greatest (and the least) impact and to make corrective interventions. (p. 14)

Clearly, this indicates a formative approach to IR. However, Volkwein also clearly outlines how summative approaches apply to IR as well, specifically IR as Scholar and Researcher in his summative example. Volkwein plainly indicates a

summative need by examples of “produces evidence so that institutional effectiveness, legal compliance, and goal attainment can be *judged* (emphasis added)” (Volkwein, 1999, p. 18).

### **Utilization Concerns**

Both professions struggle with the “use” aspect of their studies. If their studies and analyses are not being used, either immediately or in the future, neither can fulfill their effectiveness purpose. So important is the concept of use that New Directions for Program Evaluation dedicated volume 39 to the concern of utilization of evaluations (Cummings et al., 1988; Johnston, 1988; Knott, 1988; Mowbray, 1988; Muscatello, 1988; Patton, 1988; Smith, 1988). Smith considers use to be a function of design and that the design process needs improvement. Smith (1988) also provides examples of how use can be defined. For examples, observing a change implemented due to the evaluation results is as valuable as knowing why no change is implemented when the results called for it (observable versus perceptual). The “no change” action could result in a better understanding of the program by decision makers as a result of the evaluation. Many evaluators want their studies to result in immediate and major changes. Small changes can occur over time that cumulates into significant changes. As a result, incremental changes or non-immediate changes are important as immediate or holistic changes (immediate versus long-term and partial/incremental versus holistic). Finally, Smith (1988) indicates that participants can be impacted as a result of simply going through the evaluation process (process versus results).

Participants thinking differently about what they are doing can an important and long-term impact of being involved in an evaluation. Smith also provides practical and sage advice for evaluators to use program theory and stakeholder orientation to help facilitate use of results. Using program theory enables the evaluator to; understand how people and organization interact, correctly align failure (i.e., poor program implementation or right program–wrong audience), uncover unintended effects, identify program components that contribute to success, and select appropriate variables and analysis. Stakeholder orientation (involvement) entails identifying and involving the appropriate stakeholders into the evaluation. Smith contributes stakeholder involvement with results being targeted and thus the more likely results will be used. Smith also suggests building criteria into the evaluation design that judges the extent of use and ways to determine is that use is ensured. For Smith, these aspects of criteria for success, stakeholder involvement are critical components to be identified and included in the evaluation design.

Understanding how stakeholders associate with and respond to various types of reporting is one way to increase evaluation use. Knott (1988) provides a contextual understanding that can be applied to the evaluation design so that each stakeholder -physicians, hospital administrators, patients—will be more likely to use the evaluation findings. While examples are provided for helping to understanding stakeholders' backgrounds, environments, and concerns, Knott also provides suggestions for enhancing use of evaluation findings. For example,

since physicians are very autocratic, the physician will expect to make a final decision. The evaluator should provide enough information so that the physician can make a decision. In addition, physicians are use to clinical settings that employ control groups; thus, the evaluator should address concerns for not having a control group (i.e., evaluation is about establishing value and that information collected usually will suffice in adding in the decision making process). Hospital administrators are concerned with budgets and do not want to create tension between themselves and physicians. The evaluator should consider including in their proposal specific ways in which a treatment could be improved or costs reduced and find ways to garner physician support of the evaluation. Patients are concerned with confidentiality and the impact of the person who presents the consent form can have both positive and negative effects on patient participation. A thorough in-service training session about the evaluation design and reason for the study would be beneficial. Evaluators should thoroughly understand the stakeholders' backgrounds, needs, and concerns. By understanding the stakeholders better, an evaluation design can be created that enhances use of findings. Thus and similar to Smith, Knott advocates that use can be effected by the evaluation design.

Focusing so much attention on use can present concerns for the field of evaluation. Mark (2001) argues from a social better purpose that evaluations' potential is more likely to be realized if it is used to inform versus influence policies and programs. He presents three fundamental functions that must be

performed in the pursuit of social betterment: (a) what constitutes a “common good”, (b) choosing a course of action that leads to the common good, and (c) adapting that course of action. Each function requires different information. For example, determining a common good requires defining a need or problem and involves a sufficient consensus. The issue in this function is that the need or problem must first be defined and value pluralism (several opposing and valid values) is inevitable. Still, evaluation’s role in this function is to inform the determination of the common good by provide empirical evidence of the extent of the need or problem. Selecting a course of action involves assessing alternative courses of actions that can be implemented as policies or programs. However and as a result of value pluralism, it is difficult if not impossible to assign equal weights to all stakeholder groups. In other words, one groups’ values will be weighted differently than other groups’ values. In addition, evaluation results from setting may not be transferable to another setting. Evaluation’s role in this function is to provide information on the course of action that better addresses the need or problem. Finally, evaluation can provide information on the strengths and weaknesses on the adaptation of a chosen course of action. During this function, evaluation can provide systematic feedback during implementation to achieve better results. As a result of these different, but required functions for social betterment, Mark cautions that if the focus is on use, then resources can be allocated on only one of the three functions—adapting a course of action. As

a result, the other required functions get short changed in the process and evaluations' potential is not fully recognized.

Use is also a concern for IR. Ewell (1989), Kinnick (1985), and McLaughlin and McLaughlin (1989) each address concerns of use in IR from different perspectives. Ewell addresses utilization of results by providing a framework for how decision makers use information and then identifies ways to present information based upon that intended use. Kinnick identifies obstacles to and provides suggestions for enhancing the use of student outcomes information. McLaughlin and McLaughlin consider information from a managerial level activity/process perspective and provide recommendations for reducing barriers to effective use of information at each activity/process level. Ewell indicates that IR can learn from utilization evaluation researchers and take their lessons and suggestions to enhance use in their profession. Information professionals provide information to decision makers (but do not participate in decisions), are expected to supply complete and accurate information, and are expected to be value neutral. The arching implication of this is to understand how decision makers use information. Ewell elaborates on four note-worthy ways information is used: (a) to identify problems, (b) provide a context for decisions, (c) to induce action, and (d) to promote or legitimize a decision. Problem identification has to do with detecting anomalies. As a result, this implies a need for comparative information (historical trends, peer data, or expected goal measurements). He also posits that problem identification implies a need for simplicity in presentation. This

means providing a few key indicators and the presentation should clearly and easily show discrepancies. Providing a context for decisions involves linking impacts to decisions. This type of information is normally not sufficient for making a particular decision. However, it is important in understanding that the impact from a decision. For example, in trying to increase retention, a college could simply increase admission standards. However, academically prepare students tend to receive admittance offers from numerous institutions. The impact from increasing admission standards could be a decrease in the yield rate or a decrease in applications, an increase in financial aid moneys, an increase in the tuition discount rate, an increase in the demands placed upon the library, and possibly an increase in student's wanting to conduct research. Since faculty time is already committed to service, research, and teaching requirements for the college, faculty must reduce scheduled time in one or more area to support additional student research activities. Thus faculty could be impacted by the admission standards. Inducing action involves more than simply providing information. It implies supplying information in such a way that it helps decision makers come to a closure. One way to do this is to provide a probable range of variation in a result in comparison to making an alternative decision. Consider the use of confidence intervals or practical significance versus statistical significance in inducing action. Promoting or legitimizing a decision involves mobilizing support and is not about decision making. This type of information use promotes

the rationality of the action. In higher education, Ewell (1989) likens this to external accountability reporting—a growing requirement of IR.

Kinnick (1985) also considers information use with a focus on student outcomes. Kinnick groups obstacles to using into two categories: technical and organizational. Technical refers to the quality of the information. Organizational refers to characteristics of the institution's organizational and managerial setting. Organizational obstacles include the following: lack of access to information, lack of the appropriate organizational structure or framework, little or no incentive to use information, and broken links between those who develop and manage information and those of use it. Another way to view organizational obstacles is the institutional ethos that prevents information use. For example, do stakeholders know that the information exists, are departments or committees willing to work with each other and are they empowered to make suggestions or recommendations, are units required to use information in planning and budgeting, and is there a link between decision makers and information providers? Technical factors include bulky reports, data integrity, face validity, timely reporting of information, the contextual meaning of the information, and organizing reports around data and not around issues. Technical obstacles might also be viewed as producing data versus producing meaningful information. For example, cabinet or senior staff members deal with many issues on a daily basis and therefore have to effectively allocate efforts. This means that they do not have time to read through a 20-page report. As a result, executive summaries,

which tend to present concise and easily understood findings, should be considered. To enhance the use of student outcomes data, Kinnick suggests incorporating it into a specific institutional problem or process such as strategic planning such as; recruitment and retention, program reviews, comprehensive budgeting and planning, and even employing the use of mini-grants. She provides strategies for data presentation to help enhance the use of student outcomes information. While aggregate information is helpful at a holistic level for the institution, disaggregation of data at the unit level helps to make information more meaningful to departments and programs. This detailed level of information helps to unmask specific areas of concern that can be hidden in the wholeness of the data. In addition, disaggregation of data can also result in identifying areas of best practices. Using comparative formats and graphics is another suggestion of Kinnick's. Graphical and comparative reports should be arranged in such a way that stakeholders can easily glean the important information and have a sense of how the results compare to other groupings. While comparative information can likely be derived internally, use of national, aspirant, peer, Carnegie classification, or athletic division can add additional contextual meaning to information. Using short issue-specific reports for specific audiences can also help bring attention to particular timely issues or problems. Lastly, she indicates that presenting student outcomes information in combination with other information about the institution can bring greater contextual meaning to the data.

McLaughlin and McLaughlin (1989) consider overcoming barriers to effective information use by examining three levels of managerial activity/process (strategic planning, managerial control, and operational control) and how each level places constraints on information use. Strategic planning activity has a focus on overall organizational efforts. This activity is characterized by long-range time perspectives and value relationships with governance and policy. In higher education, this particular activity is more complicated with the expected roles of the faculty and the administration in governance. Managerial control activities are associated with accomplishing purposes stated within the institution's objectives. Managerial control is characterized by mid-range time perspectives and activities required to adjust to changing situations. In educational settings, this activity tends to focus on funding and managerial competency. Operational activities are more short-term focused, are high certainty activities, and focus on implementing management decisions. Next, they indicate five information support activities needed to ensure effective use: (a) selection, (b) capture, (c) manipulation, (d) delivery, and (e) influence. Selection of data to collect focuses on key questions asked by administrators or important to the institution. Selection of data is normally driven by the institution's mission, objectives, or strategies. Capturing data is data base development and administration. As a result, it involves the coding, storing, and maintaining data. This information support activity is driven by data administration (DA) or data base administration (DBA) concepts. These concepts determine how to collect

and maintain both computerized and noncomputerized data. Manipulation of data is giving creating meaningful information. This support activity identifies context, analysis, interpretation, integration, and similar activities. Information must be produced so that the user can interpret and understand it. Purpose for which the information is to be used and knowing which managerial activity/process it is to support are considered critical. Delivery requires that the information management professional (IR person) be skilled in enhancing the user's ability to integrate the information and be skilled in communicating information. Influence concerns utilization of information. A key component in utilization of information is that it should reduce uncertainty for decision makers. While McLaughlin and McLaughlin provide multiple barriers to effective information use for each of the information support activities, they also provide recommendations for enhancing effective information use at each of the managerial activity/process levels. At the strategic planning level IR members should: educate the institution about the information they can provide and the benefits in using it, gauge usefulness or reports provided and survey for untapped needs, continue to develop professionally in critical areas of communication, interpretation, and technical skills, and develop user groups or faculty advisory groups to enhance information use. At the managerial control level IR should be flexible so as to fit the resources available, keep objectives in mind when accepting new or modified activities—in other words, consider the impact of doing projects that are not part of the functions and responsibilities of the office, educate the user on which

information is best to use for the project at hand, do not compete of existing programs or services in producing information, and time is valuable—so try to accomplish multiple objectives at the same time if appropriate. IR offices can do the following at the operational level: before beginning a project scan the situation for potential problems, conduct literature searches or contact colleagues to increase knowledge about the particular project, develop a plan for the project that involves stakeholders, and keep detailed documentation and post meeting minutes/notes for all stakeholders to review.

Volume 104 of the *New Directions for Institutional Research* provides much information to the effectiveness of IR offices (Bagshaw, 1999; Knight et al., 1999; Lindquist, 1999; Muffo, 1999; Peterson, 1999; Terenzini, 1999; Volkwein, 1999; Schmidtlein, 1999). With respect to IR offices, Terenzini (1999) states, “. . . all three forms . . . are found in truly effective institutional research offices, and occasionally they are found in the same individual. More such offices and individuals are needed” (p. 29). Schmidtlein (1999) suggests that effective IR offices must recognize and work with the complexities that accompany decision making and that one way to do this is to understand organizational behavior. Bagshaw (1999) takes a different approach and applies learning theory to effectiveness. While there is ongoing interest in its effectiveness, Knight et al. (1999) indicate that empirical research is lacking in this area. Still, evaluation utilization and institutional research effectiveness is not as definitive as might be thought.

Many researchers want their studies to have an immediate and reversal impact in the operations being studied (Nagel, 1983). However, utilization is often times more complex and complicated than expected. Patton (2007) describes process use “as changes in attitude, thinking, or behavior that result from participating in an evaluation” (p. 99). As a result, understanding the use or impact of a study may be more difficult to identify. However and is evident by the dedicated issues in their respective journals, use/effectiveness are a concern for both professions.

### **Discipline**

Although there are many similarities, differences do exist between the Evaluation and IR. One obvious difference between evaluation and institutional research is that evaluation has formal teachings that lead to a degree and IR does not. Stufflebeam and Shinkfield (2007) provide a brief listing of some of the schools that offer graduate degrees in evaluation. In addition, the American Evaluation Association’s web site provides links to and information on approximately forty-five schools inside and outside of the United States of America that offer graduate degree evaluation programs (AEA, 2008b). Being a discipline helps in creating and promoting critical and different views from within the branch of learning. This internal questioning and critical review lead to evaluation’s identity crisis. The identity crisis is associated with postmodernist views (Patton, 2002; Rossi et al., 1999; Schalock, 2001; Stronach, 2001; Stufflebeam, 1998). Postmodernists cite epistemological differences when

conducting an evaluation. Thus, knowledge is constructed based upon its presuppositions, foundations, extent and validity. Therefore, knowledge can be influenced by conditions such as social and cultural upbringing, education, or political position. As a result and depending upon the perspective, there can be different and valid truths.

As for IR, there are a handful of universities that offer certificates in IR such as Florida State University, Pennsylvania State University, Ball State University, University Missouri at St. Louis, and Indiana University Bloomington. At least one university, the University of North Carolina at Greensboro, is proposing a terminal degree program in institutional research. Perhaps the humble beginnings of awarding certificates are the same process that evaluation used in its own creation of degree programs? If this is the case, then the development of an IR program leading to a degree appears to be following in the same footsteps as evaluation—albeit 20 years later.

In summary, the differences between these professions appear to center on the fact that IR is not a discipline. A discipline allows for focused topics and discussions that might not otherwise occur. How this difference surfaces between these professions is represented by each in their published studies on use/effectiveness. Evaluation researchers studied and reported on what did and did not work with respect to use. Whereas IR researchers reported on what should work with no case studies or exemplars presented. In addition, Knight et al. (1999) clearly state the need for more empirical research into the

effectiveness of institutional research offices. Since the differences appear to emanate from the discipline distinction, IR might find useful information and possible guidance by closely examining evaluation's progress and development.

### **Theory**

According to Alkin (2004), evaluation does not have any true theories. However, it does have approaches or models (Alkin, 2004, p. 4). Degree programs typically have a set of theories that undergird the discipline. Typically, theory is developed after the discipline is created rather than vice-versa. It appears that having a theory based program is a typical hallmark for entry into the formal recognition of a discipline. Recall that evaluation does have a number of degree programs. Alkin describes two general types of models:

(a) A prescriptive model, the most common type, is a set of rules, prescriptions, and prohibitions and guiding frameworks that specify what a good or proper evaluation is and how evaluation should be done; such models are exemplars; and (b) a descriptive model is a set of statements and generalizations which describes, predicts, or explains evaluation activities. Such a model is designed to offer an empirical theory. (Alkin, 2004, p. 5)

While IR employs theories derived from other disciplines, there are no known theories on institutional research.

At least on the surface, it appears that IR and Evaluation are similar in many aspects. For example, evaluators and institutional researchers share similar composition of educational backgrounds; their beginnings can be traced to the federal government; the 1960's can be considered formative years; both

appear to have a social betterment objective; both professions' utilization appears to be an ongoing concern; and it appears that both have no "true" theories.

### **Similarities: A Final Word**

When evaluation of higher education and its programs are placed in context, it tends to mirror the multiple evaluation approach of using empowerment evaluation, theory-driven evaluation, consumer based evaluation, and inclusive evaluation, postulated by Bledsoe and Graham (2005). For example and by its very nature, education is about empowerment—empowering people to become better and more productive citizens; thereby enabling society betterment. Education, as stated earlier, also tends to be theory-driven; thus, education is (or should be) very intentional about its student learning outcomes. Since a key result of education is action, knowing how those who pass through its ivory towers act is an important measure; thus, student feedback (consumer information) is an important component (such as, did students learn at a level deemed appropriate by the faculty or did their college career better prepare them for work, social and personal environments). In addition to student feedback, employer feedback (inclusive) would also be desirable. Typically this type of feedback is more difficult to acquire, requires alumni approval, and might be less valuable than originally thought. Obtaining such information helps to provide a more holistic picture of education's impact.

While process evaluation is not cited in Bledsoe and Graham, it has powerful implications because the way in which a program/service is implemented has a major impact on the delivery of services. Because IR tends to deal with existing programs and services, conducting a process evaluation is not normally a luxury that is afforded. Yet, program implementation should be considered when possible.

Although advocating any one evaluation model over another would shortchange the practical and applicable abilities of other models, and given the collegial nature of education in general, any model that involves stakeholders in a participatory role would fit the higher education environment quite well. Some evaluation models that promote and provide collegiality and collaborative efforts are; empowerment evaluation, the CIPP model, the UTOS design, theory-driven evaluation, utilization-focused evaluation, and participatory evaluation. Why would models similar to those listed be considered in IR? The reasons are clear; involving stakeholders (i.e. faculty) in the needs development stage and the methods for evaluation, there is a creation of ownership for the goals and the objectives. Without ownership there is little, if any, investment into the program. Without investment into the program, stakeholders have little incentive to investigate possible program improvements—thus ownership is critical. In summary, not only are both professions are highly contextual, but evaluation approaches appear to mirror processes that higher education might be receptive to using.

### **A Possible Theory: The Evaluation Theory Tree**

Given the volume of their similarities, can IR learn from evaluation? Just by the sheer fact that evaluation has degree programs and IR does not, evaluation is clearly more developed as a discipline than IR. As a result, might IR be able to learn from evaluation? If so, then charting the growth and development of evaluation as a profession may provide the profession of IR suggestions for charting its own disciplinary advancement. Alkin and Christie's *Evaluation Roots* (2004) may provide some focus and comparability between the two professions. The basis of this book is to correctly place evaluation researchers on one of three tree branches based upon their primary focus (methods, values, or use). However and first, the tree trunk is built on a foundation of accountability and social inquiry (social betterment). Accountability is not solely regulated to resource use, but is also designed to improve programs and society (Alkin, 2004, p. 12). For higher education, evaluations' accountability foundation sounds familiar to what many have called institutional effectiveness. Institutional effectiveness can be further delineated into the department, office, or program level. The social betterment parallel was drawn earlier. Still, the tree's roots foundation adds additional support to the similarities between evaluation and IR. It should be pointed out that the methods branch is located between the other two branches—representing a middle ground where evaluation researchers employ sound methods. Thus regardless of branch placement, the methods branch is a short reach for use and value researchers. This is an indication that

each branch theme is employed, at some level, by all evaluators. Therefore, values researchers want their evaluations used, use researchers want their expertise valued, and (hopefully) all employ sound methods. Although the evaluation theory tree is slightly altered at the conclusion of the book, the original tree provides a good basis for further discussion of and possible comparison to IR.

### **Methods Branch**

Methods do matter. As indicated above, (hopefully) all evaluation researchers employ sound methods—a foregone conclusion for any author who publishes in a peer review process. But nonetheless, any good project whether it is research, applied research, or basic reporting, requires sound methodological procedures to be considered reliable. If the methods are found to be unacceptable, even the most basic report/study can be rejected. For example, if your selection method is simply one of convenience (i.e. surveying only students who are studying in the library, studying only students in Greek organizations, interviewing only females, conducting a focus group of only two participants, or a developing a survey in isolation without a pilot testing phase) extrapolating the results to an entire population will most likely be called into question. Some of the evaluators that reside on the methods branch are Donald Campbell, Lee J. Cronbach, Peter H. Rossi, Ralph W. Tyler, and Carol H. Weiss.

IR, like evaluation, operates in a time frame that is often considered short and pressured for quick findings. As a result, some studies conducted in IR will

fall under Rossi's "good enough" rule (Alkin, 2004, p. 129). The pressure to get the project done can lead to studies that have unrecognized limitations. In these situations, IR may not be viewed as helpful as hoped for and, as a result, creditability of the office can be at stake. Under these circumstances, a strong argument can be made that expectations for the office are not appropriate (too little time was given to complete a complex/difficult assignment or the office was pushed beyond its capabilities). Still, in institutional research, as in evaluation, methods do matter.

### **Values Branch**

The values branch has the fewest identified researchers. Perhaps this is an indication of the controversial nature of that branch. Even so, the researchers placed this branch are highly respected among their peers and colleagues and, like other methods and use evaluators, their services are continually sought after indicating that statements of value, judgment, or worth are desired. Elliot W. Eisner, Ernest R. House, Michael Scriven, and Robert Stake are among values researchers.

Within IR there are divisions about injecting values into its studies and reports. As is recognized by evaluators, it is impossible to be completely value free in any study—personal beliefs and bias are a fact of life. As a result, institutional researchers will provide some level of value into their studies. This does not imply that value injecting is always a bad thing—sometimes it is necessary. For example, when unrealistic goals are given (i.e. 100% pass rate) it

is the duty of the researcher to question those values. Typically, one of two possible problems is associated with unrealistic goals. One problem could be that the bar is set so low it is impossible to not succeed. The other possible problem is the reality of failure (subjects may dropout, become ill, or lack motivation). What happens if only 95% of participants succeed, how will the stakeholders change the program so that there is a success rate of 100%? While a process or procedural change due to a single individual is possible, program (i.e. curriculum) changes usually require a pattern or trend be established—unless there are extremely high stakes involved (i.e. life or death situation). So while values are injected into IR studies, typically these are not at the same level of what some evaluators will do offering a summative statement of value or worth, thus casting judgment upon the area of study.

### **Use Branch**

Use or lack thereof is a primary concern for both professions. The use branch could be considered the most populated branch of the tree. This might be an indication on the focus to have the evaluation results used (Alkin, 2004). For example, if the evaluation results are never used, the evaluation had no impact, or no improvements were implemented, it is as if the evaluation never happened. If this is the case, then one must ask why the evaluation took place. Some identified use evaluators are Marvin C. Alkin, J. Bradley Cousins, Jean A. King, John M. Owen, Michael Q. Patton, Hallie Preskill, Daniel L. Stufflebeam, and Joseph S. Wholey.

Similar to evaluation, the use branch appears to be a primary concern for IR. This is most evident by examining mission statements from various offices. Many IR office mission statements contain wording that speaks toward providing information in a timely manner to support decision making.<sup>1</sup> Obviously this implies that use can be considered the primary branch for the IR profession too.

### **Low Hanging Fruit**

Before leaving the evaluation theory tree, there could be some low hanging fruit that can easily be taken and used as seed for institutional research. For example, IR tends to center around human behavior and there are numerous research findings within this area that provide valuable information (i.e., psychology, sociology, and economics). In particular, psychology has many applications in the study of education. Tinto's (1988) comparison of retention and suicide is a perfect example. Tinto compares the stages leading up to suicide (the ultimate departure) with the stages that students go through in their decision to leave their college (institution departure). Evaluation has also borrowed from psychology. Donaldson and Christie (2006) have written about the excellent career match between psychology majors and evaluation. Patton (1997) references empowerment evaluation as derived in part from community psychology. Finally, Scriven (2003) calls evaluation a transdiscipline. He describes transdisciplines as notable disciplines that supply tools to other

---

<sup>1</sup>Web sites visited: Elon University, UNC-Chapel Hill, University of Georgia, College of William and Mary, UNC-Greensboro, Harvard University, Stanford University, and Samford University

disciplines (i.e. psychology) while retaining their own autonomous structure and research effort.

Fincher (1985) indicates that while evaluation and IR are compatible with each other, the influence of evaluation as a role model to IR is difficult to assess. Perhaps now, given that many years have passed since Fincher's comments, is a good time to consider evaluations possible role model for IR. A good starting point of a role model of influence is an examination of a researcher from each branch of the evaluation theory tree. A brief description of their contribution to evaluation is provided along with how each of their beliefs, actions, or research transitions to IR.

Lee Cronbach, perhaps best known for Cronbach's Alpha and the generalizability theory, is clearly a methods oriented evaluator. He felt that method was always in the service of defensible inferences (Alkin, 2004). While he did not advocate a single method, he did advocate an evaluation design concept of UTOS—derived from generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Identifying four facets to which program effects could be measured is the focal point of his belief that methods matter: (a) units (i.e., populations, sites, subjects), (b) treatments (variation in treatment delivery), (c) observations (data collected—notably on outcomes), and (d) settings (i.e., clinics, hospitals, schools, worksites). Cronbach held beliefs on a number of positions that apply to many best practices cited in evaluation and institutional research studies. For example, he favored conducting many smaller studies

(triangulation) in lieu of a single grand study and he favored formative over summative evaluation (Alkin, 2004). An additional and interesting view of Cronbach's is evaluators as educators (Alkin, 2004), "The evaluator is an educator; his success is to be judged by what others have learn" (Cronbach & Associates, 1980, p. 11).

How does Cronbach fit into IR? Most obviously, sound methods are necessary to add creditability to the institutional reports and studies. Reports and studies that are weak in methods will result in distrusted and discarded efforts. Triangulation techniques are considered a best practice approach. When triangulation is not available, other techniques can substitute such as cohort, longitudinal, or cross-sectional studies. In other words, adding context to the study provides greater insight to the findings (i.e., is retention increasing, decreasing, or constant or do satisfaction surveys provide insight to current or expected retention/graduation rates). In many cases the evaluation or assessment of higher education programs leads not to answering the question of whether or not the program is working well, but rather how the program can be improved (formative versus summative evaluation).

Michael Scriven is probably best known for his position that evaluators should decide what the values are to be measured, what constitutes a good versus bad value, and that evaluators should provided a summative statement on the worth of the program. Scriven's concept of value is somewhat comparable to Consumer Reports where a "Best Buy" or "Recommended" is given to products.

Consumer Reports decides the appropriate criteria for which judgments are to be made and presents these judgments for all to see. The most recognized “consumer reports” for higher education is probably US News’ ranking of colleges. U.S. News decides what criteria to use (such graduation rates and standardized test scores) in its rankings, how much weight to assign each criterion in their analysis and provides these judgments for all to see. One key to making value judgments is identifying critical competitors so that comparisons can be made. However, Scriven goes further than most value evaluators by advocating a “goals-free evaluation.” In this type of evaluation, the program goals are rejected from the onset and it is the responsibility of the evaluator to determine which program outcomes to examine.

How do Scriven’s value evaluation concepts fit into IR? While few IR offices would advocate or want to engage in a goals-free approach, values are incorporated in many IR studies and reports in the form of comparisons or benchmarks. For example, many studies incorporate comparisons to national averages, Carnegie classifications, public/private college control, peer and aspirant schools, or athletic conference membership results. Once again, placing the study in greater context adds to the richness of the results and interpretations. For example, knowing that a schools standardized test scores, average faculty salaries, endowment, retention and graduation rates, and overall enrollment are all significantly above peer schools, might inform cabinet staff that it is time for a new set of peer schools.

Along with Guba, Daniel Stufflebeam developed the CIPP model (context, input, process, and produce). While the CIPP model has a very “methods” feel to it, collegiality is major component to the model that helps promote use of the evaluation results. For example, a representative stakeholder panel is formed and it is through this panel of representatives that questions, plans, reviewing draft reports, and dissemination of information are brainstormed, developed, agreed upon, and shared (Stufflebeam, 2000). The collegiality creates buy-in and momentum for the evaluation results to be used.

Stufflebeam’s collegiality emphasis obviously fits nicely into higher education. However, how does it fit into IR? The take-away point from Stufflebeam’s CIPP model is finding and engaging in the tactics that increase probabilities of results being used. This means understanding what type of reports and analysis to produce for various decision makers. As an example, a particular vice president may like volumes of data and another may want an executive summary. It will be the responsibility of the IR officer to understand and use effective reporting means. There are a number of ways to find out what is effective by reviewing previous reports and having discussions with co-workers, faculty, and staff members that will provide insight into the reporting formats. Through collaborative efforts, the appropriate styles of reporting can be tailored to each decision maker and stakeholder involvement will create the support required to increase using the results. One critical point about use should be understood. Just because the IR office is unaware of the use of the study/report

does not mean it is not having an impact. The IR office should follow the distribution chains and ask if the receivers are finding the studies/reports useful. Soliciting this type of feedback might provide additional insight into what is and what is not working with respect to use.

### **Reality**

While theory tends to yield idealism, most of us work in the world of messy reality. Politics, time constraints, limited resources, and unrealistic expectations run rampant during life. Humans are complex and difficult to work with. While the ends might justify the means, the means may not follow what was originally planned. For example, a journalist interviewing the boxer who won the fight asked, “Didn’t you have a plan before entering the ring?” The boxer replied, “Yea—but then the first punch was thrown.” The poignant punch is that sometimes plans change—because reality enters the picture. The reality in which IR operates is: higher education is not a high stakes situation (life or death); our accountability stakeholders, while many, tend to be politicians; collaborative based studies tend to be the most applicable within higher education; theory-driven evaluations, like process evaluations, are not usually considered; and inclusive evaluations, at least at the employer level, entail some assumptions about the employer’s ability to accurately evaluate the specific worker/alumni.

The examination of the evaluation theory tree and how the methods, values, and use branches relates to IR, as well as a brief overview of selected evaluators and how their focus on evaluation also relates to IR, provides

additional examples of likeness between the two professions. The similar realities and contextual settings that both operate in add additional resemblances.

### **Summary**

While IR and evaluation have much in common, evaluation appears to have cleared a path that IR can follow. For example, if IR wants to become recognized as a discipline within education, it might want to follow the prescriptive and descriptive modeling identified by Alkin. Through teaching of the discipline, IR can become streamlined—that being a more common set of functions across institutions rather than its current and vast institutional contextual settings for determining office functions. Once IR develops models, offices will have more common processes and procedures. While administrative offices may never be fully accepted as colleagues to teaching faculty, IR credentials will at least be more in line with the faculty by having a theory-based or model/approach based discipline.

The Southern Association of Colleges and Schools may have provided an opening for IR to become more connected to evaluation through its new criteria (Quality Enhancement Plan—QEP). Still, the QEP might prove to be a compelling process that helps evaluation to take a stronger stance within the higher education community. In one sense, the QEP could be to evaluation in higher education to what the war on poverty was to evaluation in federal programs.

Just as Alkin and Christie placed evaluators on each branch of the evaluation theory tree based upon their emphasis, perhaps IR offices will consider which branch they would place themselves upon an IR tree? At the same time, perhaps IR offices may recognize which branch they would want to be placed and work towards making that happen. Is a model or theory for IR lurking in the evaluation theory tree? Can IR improve by asking three simple questions: has IR added appropriate value to our studies and reports, what has IR done to enhance the use of results, and is IR paying enough attention to methods? A comparison to both professions having a purpose of institutional (program) effectiveness has already been made. Is a model or theory yet to be derived from evaluation?

If IR and evaluation are two sides of the same coin, where does this leave IR? Are evaluation models IR models too? Could IR become a transdiscipline as Scriven describes or will IR be impacted by the transdiscipline of evaluation? Enough evidence has been presented that can easily lead one to believe that IR and evaluation are opposite sides of the same coin. However and if this is the case, why haven't more evaluation models been used in IR? Could it be that evaluation models have not been clearly linked to IR? Is it possible that an IR office is in fact doing evaluation, but an evaluation office is not necessarily doing IR? While the reasons for lack of modeling use could continue, perhaps the best reason is that IR, quite plainly, does not have a model of its own. Perhaps the establishment of an IR model will be the missing piece to the puzzle?

The following methodology will be used to answer five research questions:

1. Does an Evaluation Theory Tree (Methods, Use, and Values branches) fit Institutional Research?
2. How do “Evaluators” differ from “Non-evaluators” in their placement on each branch of the Evaluation Theory Tree?
3. Which approaches to conducting IR are significantly related to branches of Core IR reporting?
4. Is there a relationship between office FTE and branch scores?
5. Is there a relationship between institution type and branch scores?

## CHAPTER III

### METHODS

#### Survey Administration

In this study an anonymous survey was given to Directors of Institutional Research. Literature review guided construction of the survey instrument. In particular, Christie's (2001) evaluation theory-to-practice survey (Evaluation Theory Tree) instrument was used to help guide the creation of the methods, use, and values domains for this survey. Individuals who are identified as Directors of Institutional Research were contacted to participate in the study. These people were chosen due to their ability to answer questions that relate to working with stakeholders, decision makers, and knowledge of use of the reports/studies produced by their office. The instrument was designed to be distributed electronically. Email and title information were obtained from the Higher Education Publications, Inc (HEP). The HEP directory results in 2,031 email addresses of Directors of Institutional Research. The instrument was developed and pilot tested with 200 randomly selected directors. Two emails were sent resulting in 67 responses (or 33% response rate) for the pilot test. Reliability of the pilot test resulted in an overall Cronbach's Alpha of 0.89. Slight modifications were made to the survey. The final version of the survey was distributed/emailed to 1,831 individuals. Following the protocol of the pilot test,

two emails notifications for participation were sent two weeks apart. One-hundred thirty-one email addresses were returned as bad addresses. The survey nets 573 participants, or a 34% valid response rate (573/1,700).

The final version of the survey (see Appendix A) contains three primary sections: (a) Core IR Reporting, (b) Assessment/Evaluation Practices, and (c) Approaches to Conducting IR. The Office Reporting Practices section contains 13 items; the Assessment/Evaluation Activities contains 20 items; and 16 items are in the Approaches to Conducting IR. Each primary section contains items that are identified with methods, use, and values domains. The survey contained two secondary sections: (a) Helping Stakeholders Conduct Assessment/Evaluation (contains six items) and (b) Practices in Developing Instruments (contains six items). There were two branching questions; one for entry in to Assessment/Evaluation Activities and Approaches (Approximately how many assessment or evaluation related activities does your office conduct during a typical academic year?) and one for entry into Practices in Developing Instruments (Approximately how many measurement instruments (i.e., surveys, tests, rubrics) does your office help to create during an academic year?). Each branching question provided a range for selecting volume of activities in conducting assessment activities and creation of measurement instruments. Selection choices are the same for branching questions (None, 1-4, 5-8, 9-12, 13-16, 17-20, and 21 or more). Not answering the branching question or selecting a response of 'None' allowed the subject to skip that entire section. The

survey ended with 14 demographic questions and one open text box for additional comments. Demographic questions include gender, race/ethnicity, age, educational background, degree obtainment, years in IR, years as a Director, relationship to IR, IR knowledge, College teaching, IR Office FTE, selectivity of current employer, description of employer (i.e., 4-year public institution), and American Evaluation Association member.

The Core IR Reporting section contained the following items arranged in order of their respective branches of Use, Methods, and Values. Participants are asked rate each item's similarity to their actual office practices when conducting or producing reports. Likert-type scale selection choices are 1=None, 2=Very little, 3=Some, 4=Quite a bit, and 5=Very much.

#### *Values Branch*

- My office regularly includes the main people using the reports to help interpret the meaning of the data.
- Stakeholders help to judge the quality of the methods employed.
- Stakeholders help to judge the appropriateness of the methods employed.

#### *Methods Branch*

- Recommendations are regularly included in our reports.
- My office regularly provides interpretations of the results.
- My office regularly provides judgments in reports.
- Suggestions are regularly included in our reports.
- Discussions are regularly included in our reports.
- Methodological procedures are regularly included in reports.

#### *Use Branch*

- Unless it is highly sensitive information, the reports are available for anyone to see.
- Benchmark or comparison figures are regularly included

- Limitations are regularly included in our reports
- My office seeks feedback on our reports (i.e., the usefulness of or needed improvements to).

The Assessment/Evaluation Practices section contained the following items arranged in their respective branches of Use, Methods, and Values.

Participants were asked to rate each item's similarity to their actual assessment/evaluation practice. Likert scale selection choices are 1=None, 2=Very little, 3=Some, 4=Quite a bit, and 5=Very much.

#### *Values Branch*

- The main people using the results always help interpret the meaning of the data.
- Appropriate stakeholders judge the quality of the research methods employed.
- Primary stakeholders' assumptions are integrated into the study process.
- Only the primary stakeholders are involved in the process.

#### *Methods Branch*

- I tend to use a mixture of quantitative and qualitative methods.
- Recommendations are included in the study's report.
- Interpretations of meaningfulness of results are provided in the study's report.
- Judgments of meaningfulness of results are provided in the study's report.
- Discussions are included in the study's report.

#### *Use Branch*

- Contextual information about the results is included in the study's report.
- A primary role for my office is to assess the study's quality against acceptable norms for conducting research.
- The process is guided by a conceptual framework, model, or theory.
- Different versions of the same report are produced so that different stakeholders understand the findings.
- Only instruments with evidence of reliability or validity are used (these include local or national instruments).

- My practice reflects a primary purpose of IR that enhances knowledge and strategies for designing and implementing improvements to programs and services, and not just to assess effectiveness.

The Approaches to Conducting IR section contained the following items arranged in their respective branches of Use, Methods, and Values. Participants were asked to rate each items' similarity to their actual approach to practicing IR. Likert scale selection choices are 1=None, 2=Very little, 3=Some, 4=Quite a bit, and 5=Very much.

#### *Values Branch*

- Stakeholders participate in the collection of the data.
- Stakeholders participate in the analysis of the data.
- Stakeholders participate in the interpretation of the data.

#### *Methods Branch*

- Interpretations are included.
- Judgments are included.
- I am concerned that other independent researchers can replicate our studies.

#### *Use Branch*

- IR works with the main people using the results to help determine the next steps.
- IR creates changes in the culture of the institution.
- I am focused on the involvement of decision-makers versus the other stakeholders.
- My office builds upon the current generalized knowledge base of the particular program being studied.
- Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.

Each section was evaluated for items that were contradictory or considered highly related to other items for possible exclusion to branches. Core IR Reporting had no items identified for exclusion. The Assessment/Evaluation Practices section resulted in five items removed: (a) Only decision-makers are involved in the process, (b) Appropriate stakeholders judge the appropriateness of the research methods employed, (c) My primary method of choice is quantitative, (d) All relevant stakeholders are involved in the process, and (e) My primary method of choice is qualitative. The Approaches to Conducting IR section resulted in three items removed: (a) It is more important to provide decision-makers what they need versus other stakeholders, (b) I am concerned that the results from our studies are useful for other populations or other academic settings, and (c) Stakeholders are included in the study's process (i.e., collection, analysis, or interpretation of the data).

The following methodology was used to answer the five research questions:

1. Does an Evaluation Theory Tree (Methods, Use, and Values branches) fit Institutional Research?
2. How do “Evaluators” differ from “Non-evaluators” in their placement on each branch of the Evaluation Theory Tree?
3. Which branches on Approaches to Conducting IR tree are significantly related to Core IR reporting?
4. Is there a relationship between office FTE and branch scores?

5. Is there a relationship between institution type and branch scores?

### **Theory Fit: Confirmatory Factor Analysis**

Structural Equation Modeling (SEM) for confirmatory factor analysis (CFA) was used to test Research Question 1 (Does the Evaluation Theory Tree fit Institutional Research?). SEM was selected as the statistical methodology to answer this question because of its several advantages over regression modeling, including its use of confirmatory factor analysis to reduce measurement error by having multiple indicators per latent variable, the desirability of testing models overall rather than coefficients individually, the ability to test models with multiple dependents, and the desirability of its strategy of comparing alternative models to assess relative model fit. AMOS (Arbuckle, 2006) was used to conduct the CFA analysis. The analysis was conducted using maximum likelihood estimation. Data were screened for approximate multivariate normality using visual inspection of QQ plots (results are in Appendix B). Q-Q plots indicated that the data was approximately univariate normal and provided no indication of a lack of multivariate normality.

Three possible models were tested for each primary section of the survey: (a) Office Reporting Practices, (b) Assessment/Evaluation Practices, and (c) Approaches in Conducting IR. Testing each section of the survey separately helped to avoid missing data issues introduced by the branching structure of the survey. That is, participants could skip sections of the survey by selecting “None” on a branching question. The implication is that the respondent could not answer

that section because she/he did not engage in that activity (e.g., assessment). As a result, participants could skip entire sections of the survey. Taking that missing data into account, the sample size would not support stable estimation of all survey items at once. The proposed tree models support nested models. As a result, it was necessary to evaluate the fit for a one branch, two branch, or three branch best fitting model. Whereas AMOS provided a number of fit statistics, the chi-square (CMIN), the Root Mean Square Error of Approximation (RMSEA), and the Comparative Fit Index (CFI) were the primary measures for assessing model fit. A RMSEA less than 0.08 and a CFI greater than 0.90 are considered indicators of a good fitting model (Raykov & Marcoulides, 2006; Hoyle, 1995). In addition and because these are nested models, a chi-square difference test was used when comparing the one, two, and three branch models (Hoyle, 1995). The three branch models had items that were separated by methods, use, and values; two branch models collapsed methods and uses branches into a single branch and retained the values branch; one branch models collapsed all items into a single branch. The decision to collapse the methods and use branches for the two branch model was based upon how these two branches are related to each other; that being, decision-makers use of analysis was based upon sound or acceptable methods employed.

### **Branch Placement: Multivariate Analysis of Variance and Descriptive Discriminant Analysis**

Multivariate analysis of variance (MANOVA) was used to test Research Question 2 (How do “Evaluators” differ from “Non-evaluators” in their placement

on each branch of the Evaluation Theory Tree?). Evaluators were defined as institutional researchers that have a degree related to evaluation or membership (past or present) in the American Evaluation Association (AEA). An Evaluation Research Methods degree is considered an evaluation related discipline and is a selection choice for highest degree held. The concept behind this definition is that this group has had a greater exposure to more formal evaluation methods and practices than non-evaluators. Whereas organizations such as the Association for Institutional Research (AIR) might offer workshops and other types of training in assessment, a degree related to evaluation or membership in AEA is believed to provide more direct evidence of exposure to such methods and practices.

Mean scale scores were created to determine branch placement differences between Evaluators and Non-Evaluators. Mean scale scores were calculated using a respondents' average response to that branch area (for example, the average Value score for Core IR Reporting). Using mean scale scores in this study was appropriate due to the consistency in the five-point likert response scales across domains (DiStefano, Zhu, & Mindrila, 2009).

During the collection of information from voluntary surveys (which was the case in this study), subjects often skipped questions, thus resulting in missing data. Preliminary analyzes of the data indicated that no item has missing values exceeding three percent. Tabachnick and Fidell (2001) state that "if only a few data points, say, 5% or less, are missing in a random pattern from a large data

set, the problems are less serious and almost any procedure for handling missing values yields similar results” (p. 59). As a result, missing data were addressed by substituting the mean score for that item into the missing value.

Discriminant analysis was used to help describe statistically significant differences between Evaluators and Non-evaluators through the discriminant functions. Discriminant analysis has two very nice features: parsimony of description and clarity of interpretation. Thus discriminant analysis will help to describe group differences. The description of which mean scale scores contribute to group differences may provided insight into how formal evaluation training impacted IR practice.

#### **Approaches Branches and Core Reporting: Multiple Linear Regression**

Whereas specific office functions and responsibilities varied across institutions, reporting is a necessary and required task for each institutional research office; thus, it can be viewed as a common domain. Although evidence to date has not been presented indicating best practices in Core IR reporting, the items incorporated in the Methods, Use, and Values branches of that section of the survey arguably could be viewed as such. Best practice could include the inclusion of methods, benchmarks, recommendations, limitations, suggestions, discussions, interpretations, and judgments into reports; as well as collaboration with various stakeholders in the creation of reports, the appropriate dissemination of reports, and seeking feedback for reporting improvements. If this is the case, then higher mean scores in the Core IR Reporting section are an indication of

greater application of these practices. As a result, mean scores were used as a proxy for measuring best practices.

Multiple linear regression (MLR) was used to test Research Question 3 (Which approaches to conducting IR are significantly related to Core IR Reporting branch mean scores?). This analysis was exploratory in nature and was intended to provide guidance to enhancing skill sets. The use of MLR was the preferred analysis due to its ability to estimate the conditional effects of several independent variables on a dependent variable. Core IR Reporting mean branch scores were regressed onto the items in Approaches to Conducting IR. The standardized beta weights for statistically significant branches were the focus for additional investigation(s). Evaluating standardized beta weights allowed for a comparable comparison of the predictive capability of the independent variables. Scales were standardized by setting the means to 0 and standard deviations to 1. As a result, all measurements were rescaled to a common unit of measure.

#### **Relationship between Branch Score and Office FTE: Scatter Diagrams**

This question was intended to be exploratory in nature. Does office staffing, particularly Full Time Equivalent (FTE) (as defined by the participant), have a relationship with branch scores? Office staffing FTE could possibly play a role in how participants responded to various questions. Whereas a relationship was possible, evidence presented in the literature review indicated that the IR operation and functions appear to be institutionally driven. In other

words, how IR operates at any given institution is a function of that institutions needs. Still, analyzing this question helped to clarify IR's unique institutionally contextually driven functions, operations, and reporting. The findings could help in understanding staffing needs for IR offices.

### **Relationship between Branch Score and School Type: Box Plots**

This question was also intended to be exploratory in nature. Does school type, particularly non-profit 2-year and 4-year schools, have a relationship with branch scores? School type could possible play a role in how participants responded to various questions. Although a relationship was possible, evidence presented in the literature review indicated that the IR operation and functions appeared to be institutionally driven. In other words, how IR operates at any given institution was a function of that institution's needs. Still, analyzing this question helped to clarify IR's unique institutionally contextually driven functions, operations, and reporting. Findings could help to understand how school type plays a role in shaping the functions, operations, and reporting of IR offices. This in turn could be used to guide potential employees towards institutions for a better employment fit.

## CHAPTER IV

### RESULTS

This chapter presents the results of the current study. All data from this survey was self-reported. Descriptive analyses were conducted on participants' responses. The results illustrated the demographic, educational, and professional background of participants. Confirmatory Factor Analysis (CFA) was used to test the fit for one, two, and three-branch models of the Evaluations' Tree Theory to institutional research in 3 survey areas: (a) Core IR Reporting, (b) Assessment/Evaluation Practices, and (c) Approaches to Conducting IR. Multivariate Analysis of Variance (MANOVA) was used to test differences between evaluators and non-evaluators; significant differences would be further analyzed with Discriminate Analysis. Multiple Linear Regression (MLR) was used to test various Approaches to Conducting IR that are related to the mean overall Core IR Reporting score.

#### **Demographic Information**

Twenty-one participants who did not identify themselves directors of IR were removed from all analysis. As indicated in Table 1, most participants were female (53.5%), white (89.3%), 51 years of age or older (51.0%), and hold a doctorate degree (51.2%). When considering work related experience and knowledge (Table 2), 115 (23.4%) participants have been in IR for 4-years-or-

less, 211 (38.9%) have been a director for 4-years-or-less, most participants rate their knowledge of IR as either very good (37.8%) or good (37.6%), and most work fulltime in IR (75.9%). A cross-tab analysis (Table 3) indicates that 124 (59.6%) of the 208 4-years-or-less directors have been in IR for 4 years or less.

**Table 1**  
***Demographic Description of Participants***

<b>Demographic</b>	<b><i>n</i></b>	<b>%</b>
<u>Gender:</u>		
Female	288	53.5
Male	250	46.5
<u>Race/Ethnic:</u>		
American Indian or Alaska Native	3	0.6
Asian	14	2.6
Black or African American	20	3.8
Hispanic/Latino	20	3.8
White	476	89.3
<u>Age:</u>		
Less than 30	10	1.9
30 to 35	37	7.0
36 to 40	52	9.8
41 to 45	80	15.1
46 to 50	80	15.1
51 to 55	97	18.3
56 or older	173	32.7
<u>Highest Degree Obtained:</u>		
Some college	2	0.4
Bachelors	33	6.1
Masters	227	42.3
Doctorate	275	51.2

**Table 2*****Work Experience and Knowledge***

<b>Demographic</b>	<b><i>n</i></b>	<b>%</b>
<u>Years conducting IR:</u>		
Less than 1	6	1.1
1 to 4	119	22.3
5 to 8	105	19.7
9 to 12	106	19.9
13 to 16	71	13.3
17 to 20	46	8.6
21 or more	80	15.0
<u>Years as Director of IR:</u>		
Less than 1	14	2.6
1 to 4	197	36.3
5 to 8	115	21.2
9 to 12	96	17.7
13 to 16	54	10.0
17 to 20	25	4.6
21 or more	41	7.6
<u>IR Knowledge:</u>		
Very good	186	37.8
Good	185	37.6
Average	101	20.5
Fair	17	3.5
Poor	3	0.6
<u>Relationship to IR:</u>		
IR is my full-time job	401	75.9
IR is a part of my full-time job	122	23.1
IR is a part-time job	5	0.9

**Table 3*****Cross Tabulation: Director of IR and Years in IR***

		How many years have you been conducting institutional research?							Total
		< 1	1-4	5-8	9-12	13-16	17-20	21+	
How many years have you been a director (or higher level) of institutional research?	< 1	5	6	1	1	1	0	0	14
	1-4	1	112	41	23	10	3	4	194
	5-8	0	0	63	26	15	4	4	112
	9-12	0	1	0	55	25	8	7	96
	13-16	0	0	0	1	20	18	14	53
	17-20	0	0	0	0	0	12	13	25
	21+	0	0	0	0	0	1	38	39
Total		6	119	105	106	71	46	80	533

Two hundred twenty-two participants describe their schools as Private not-for-profit (42.0%), followed next by 159 (29.4%) Public 2-year, and 130 (24.1%) Public four-year institutions. Fifty-three (9.8%) identified themselves as either having an evaluation research methods degree or having membership in the American Evaluation Association. Write-in highest degree held (Other) accounted for 10.7% of the sample. When possible these degrees were coded to existing degree choices. After the recoding of write-in degrees, selected highest degree holdings are Education (23.4%), Psychology (13.1%), Business Administration (10.9%), Sociology (8.1%), Higher Education (6.9%), Math (3.4%), and Public Administration (3.4%). Table 4 provides more specific details of highest degree held.

**Table 4*****Highest Degree of Respondents***

<b>Degree</b>	<b><i>n</i></b>	<b>%</b>
Anthropology	7	1.3
Art	3	0.6
Biology	7	1.3
Business Administration	58	10.9
Computer Science	10	1.9
Economics	8	1.5
Education	125	23.4
English	7	1.3
Evaluation Research Methods	20	3.7
Higher Education	37	6.9
History	8	1.5
Human Services	5	0.9
Information Science	5	0.9
Mathematics	18	3.4
Music	3	0.6
Nursing/Medicine	4	0.7
Political Science	11	2.1
Public Administration	18	3.4
Psychology	70	13.1
Public Health	3	0.6
Public Policy	10	1.9
Religion	5	0.9
School Administration	5	0.9
Sociology	43	8.1
Statistics	12	2.2

Tables 5 through 7 provide descriptive statistics on the items associated with each section. The tables provide information on each item that includes the total number of responses, minimum and maximum ranges of scale selection, the means, and standard deviations.

**Table 5**

***CORE IR Reporting Descriptive Statistics***

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Methodological procedures are regularly included in reports.	573	1.00	5.00	3.42	1.07
Benchmark or comparison figures are regularly included.	573	1.00	5.00	3.69	.89
Recommendations are regularly included in our reports.	573	1.00	5.00	3.11	1.11
My office regularly includes the main people using the reports to help interpret the meaning of the data.	573	1.00	5.00	3.43	1.01
Unless it is highly sensitive information, the reports are available for anyone to see.	573	1.00	5.00	3.95	1.04
Stakeholders help to judge the appropriateness of the methods employed.	573	1.00	5.00	2.90	1.02
Stakeholders help to judge the quality of the methods employed.	573	1.00	5.00	2.77	1.03
My office regularly provides interpretations of the results.	573	1.00	5.00	3.89	.96
My office regularly provides judgments in reports.	573	1.00	5.00	2.85	1.10
My office seeks feedback on our reports (i.e., the usefulness of or needed improvements to).	573	1.00	5.00	3.73	1.00
Limitations are regularly included in our reports.	573	1.00	5.00	3.56	1.04
Suggestions are regularly included in our reports.	573	1.00	5.00	3.19	1.11
Discussions are regularly included in our reports.	573	1.00	5.00	2.97	1.11

**Table 6*****Assessment/Evaluation Practices Descriptive Statistics***

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
The main people using the results always help interpret the meaning of the data.	547	1.00	5.00	3.46	.92
Appropriate stakeholders judge the quality of the research methods employed.	547	1.00	5.00	2.70	.92
Primary stakeholders' assumptions are integrated into the study process.	547	1.00	5.00	3.26	.84
Only the primary stakeholders are involved in the process.	547	1.00	5.00	2.97	.84
I tend to use a mixture of quantitative and qualitative methods.	547	1.00	5.00	3.30	1.07
Recommendations are included in the study's report.	547	1.00	5.00	3.26	1.02
Interpretations of meaningfulness of results are provided in the study's report.	547	1.00	5.00	3.59	.94
Judgments of meaningfulness of results are provided in the study's report.	547	1.00	5.00	3.30	.98
Discussions are included in the study's report.	547	1.00	5.00	3.01	1.00
Contextual information about the results is included in the study's report.	547	1.00	5.00	3.62	.81
A primary role for my office is to assess the study's quality against acceptable norms for conducting research.	547	1.00	5.00	3.19	1.07
The process is guided by a conceptual framework, model, or theory.	547	1.00	5.00	3.33	.96
Different versions of the same report are produced so that different stakeholders understand the findings.	547	1.00	5.00	2.47	1.08
Only instruments with evidence of reliability or validity are used (these include local or national instruments)	547	1.00	5.00	3.28	.89
My practice reflects a primary purpose of IR that enhances knowledge and strategies for designing and implementing improvements to programs and services, and not just to assess effectiveness.	547	1.00	5.00	3.66	.97

**Table 7*****Approaches to Conducting IR Descriptive Statistics***

	<b><i>N</i></b>	<b><i>Min</i></b>	<b><i>Max</i></b>	<b><i>M</i></b>	<b><i>SD</i></b>
Stakeholders participate in the collection of the data.	291	1.00	5.00	2.94	1.00
Stakeholders participate in the analysis of the data.	291	1.00	5.00	2.65	1.02
Stakeholders participate in the interpretation of the data.	291	1.00	5.00	3.32	.96
Interpretations are included.	291	1.00	5.00	3.47	.95
Judgments are included.	291	1.00	5.00	2.71	.98
I am concerned that other independent researchers can replicate our studies.	291	1.00	5.00	2.44	1.06
IR works with the main people using the results to help determine the next steps.	291	1.00	5.00	3.76	.84
IR creates changes in the culture of the institution.	291	1.00	5.00	3.25	.92
I am focused on the involvement of decision-makers versus the other stakeholders.	291	1.00	5.00	3.31	.88
My office builds upon the current generalized knowledge base of the particular program being studied.	291	1.00	5.00	3.54	.79
Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.	291	1.00	5.00	3.22	1.04

### Confirmatory Factor Analysis Results

Reliability analysis for each branch on each tree was performed. Table 8 presents the coefficient alpha results of that analysis. Reliability was low on the following areas: the Use branch of the Core IR reporting tree, the Values branch of the Assessment/Evaluation Practices tree, and the Methods and Use branch of the Approaches to Conducting IR tree. Reliability measurements this low could have impacted the power to detect differences should any exist.

**Table 8**

***Coefficient Alpha***

Tree	Branches		
	Values	Methods	Use
Core IR Reporting	0.79	0.85	0.52
Assessment/ Evaluation Practices	0.61	0.78	0.68
Approaches to Conducting IR	0.73	0.63	0.64

AMOS v17 was used for CFA analyses (Arbuckle, 2006). The 13 Core IR Reporting items were explicitly constrained to load on only one branch in multiple branch models (i.e., follow simple structure). In the 3-branch model, three items only loaded on Values, five items loaded only on Methods, and five items loaded only on Uses; in the 2-branch model 10 items loaded only on Methods plus Uses, and three items loaded only on Values; in the 1-branch model all items loaded on the Methods plus Uses plus Values. Table 9 summarizes results for each model tested.

**Table 9*****CORE IR Reporting Model Results***

	$\chi^2$	<i>df</i>	<i>p</i> -value	CFI	RMSEA
3-Branch Model (Values, Methods, Uses)	287.16	62	$p < .001$	.92	.08
2-Branch Model (Values, Methods + Uses)	342.11	64	$p < .001$	.90	.09
1-Branch Model (Values + Methods + Uses)	998.34	65	$p < .001$	.67	.16

As indicated in Table 9, the 3-branch model was the only model tested that meets the fit criterion of CFI greater than or equal to 0.90 and RMSEA less than or equal to 0.08. Whereas the 2-branch model was close to the criterion for fit, a  $\chi^2$  difference test indicated that the 3-branch model was statistically better fitting ( $\Delta\chi^2 = 54.95$ ,  $\Delta df = 2$ ,  $p < 0.01$ ).

Table 10 indicates the estimates, standardized estimates, and standard errors for each item in the 3-branch model. Table 11 indicates the covariance among branches (factors). As indicated in Table 11, factor covariances (which can be interpreted as correlations) are statistically significant at the 0.001 level for both models. Although these associations are significant, they are sufficiently low to provide discriminant validity evidence for the different constructs. Figure 1 provides the path diagram for the 3-branch model.

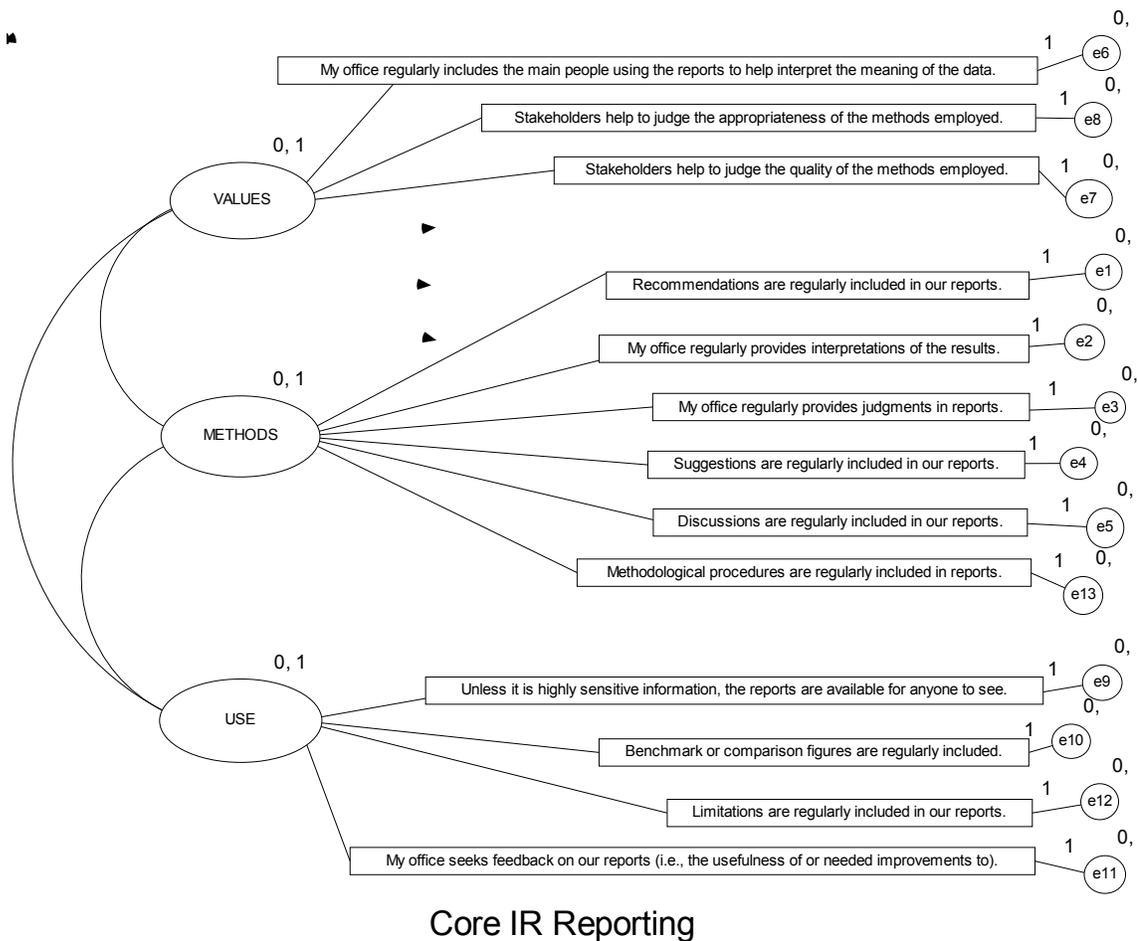
**Table 10*****Estimates, Standardized Estimates, and Standard Errors for the 3-Branch Model***

	<b>3-Branch Model</b>		
	<b>Estimate</b>	<b>Standard Estimate</b>	<b>Standard Error</b>
Benchmarks	0.344 <sup>***</sup>	.388	.042
Discussions	0.794 <sup>***</sup>	.714	.042
Feedback	0.612 <sup>***</sup>	.612	.046
Interpretations	0.640 <sup>***</sup>	.669	.037
Judgments	0.719 <sup>***</sup>	.654	.043
Limitations	0.630 <sup>***</sup>	.606	.048
Main People	0.474 <sup>***</sup>	.468	.042
Methods	0.481 <sup>***</sup>	.448	.045
Recommendations	0.894 <sup>***</sup>	.807	.040
Sensitive	0.281 <sup>***</sup>	.269	.050
Stake Appropriate	0.925 <sup>***</sup>	.909	.037
Stake Quality	0.947 <sup>***</sup>	.922	.037
Suggestions	0.970 <sup>***</sup>	.874	.038

Significant at  $p < 0.001$ **Table 11*****Covariance among Branches (Factors)***

<b>3-Branch Model</b>			<b>Covariance</b>	<b>Standard Error</b>
Use	↔	Values	.474 <sup>***</sup>	.046
Methods	↔	Values	.358 <sup>***</sup>	.041
Use	↔	Methods	.749 <sup>***</sup>	.035

Significant at  $p < 0.001$



**Figure 1. Path Diagram for the 3-Branch Model (CORE IR Reporting)**

The 15 Assessment/Evaluation Practice items were explicitly constrained to load on only one branch in multiple branch models. In the 3-branch model four items loaded only on Values, five items loaded only on Methods, and six items loaded only on Uses; in the 2-branch model 11 items loaded only on Methods plus Uses, and four items loaded only on Values; in the 1-branch model all items loaded on the Methods plus Uses plus Values. Table 12 summarizes results for each model tested.

**Table 12*****Assessment/Evaluation Model Results***

	$\chi^2$	df	p-value	CFI	RMSEA
3-Branch Model (Values, Methods, Uses)	287.51	87	$p < .001$	.90	.07
2-Branch Model (Values, Methods + Uses)	374.37	89	$p < .001$	.86	.08
1-Branch Model (Values + Methods + Uses)	466.16	90	$p < .001$	.82	.09

As indicated in Table 12, the 3-branch model was the only model tested that meets the fit criterion of CFI greater than or equal to 0.90 and RMSEA less than or equal to 0.08. While the 2-branch model was not as close to the criterion for fit as in Core IR Reporting, a  $\chi^2$  difference test was conducted. The results indicated that the 3-branch model was statistically better fitting ( $\Delta\chi^2 = 86.86$ ,  $\Delta df = 2$ ,  $p < 0.01$ ). In Table 9, the estimates, standardized estimates, and standard errors for each item in the 3-branch model are displayed.

In Table 13 the estimates, standardized estimates, and standard errors for each item in the 3 and 2 branch models are shown and in Table 14 the covariance among branches (factors) are displayed.

As indicated in Table 14, factor covariance (that can be interpreted as correlations) were statistically significant at the 0.001 level for both models. Figure 2 provides the path diagram for the 3-branch model. The covariance

between Methods and Use is considered high indicating that these factors may be measuring the same or highly related constructs. Concerns over the high association is ameliorated (somewhat) by the finding that the 3-branch model fit better. The covariance between Values and Use, and Methods and Values was moderate.

**Table 13**

***Estimates, Standardized Estimates, and Standard Errors for Each Item in the 3-Branch Model***

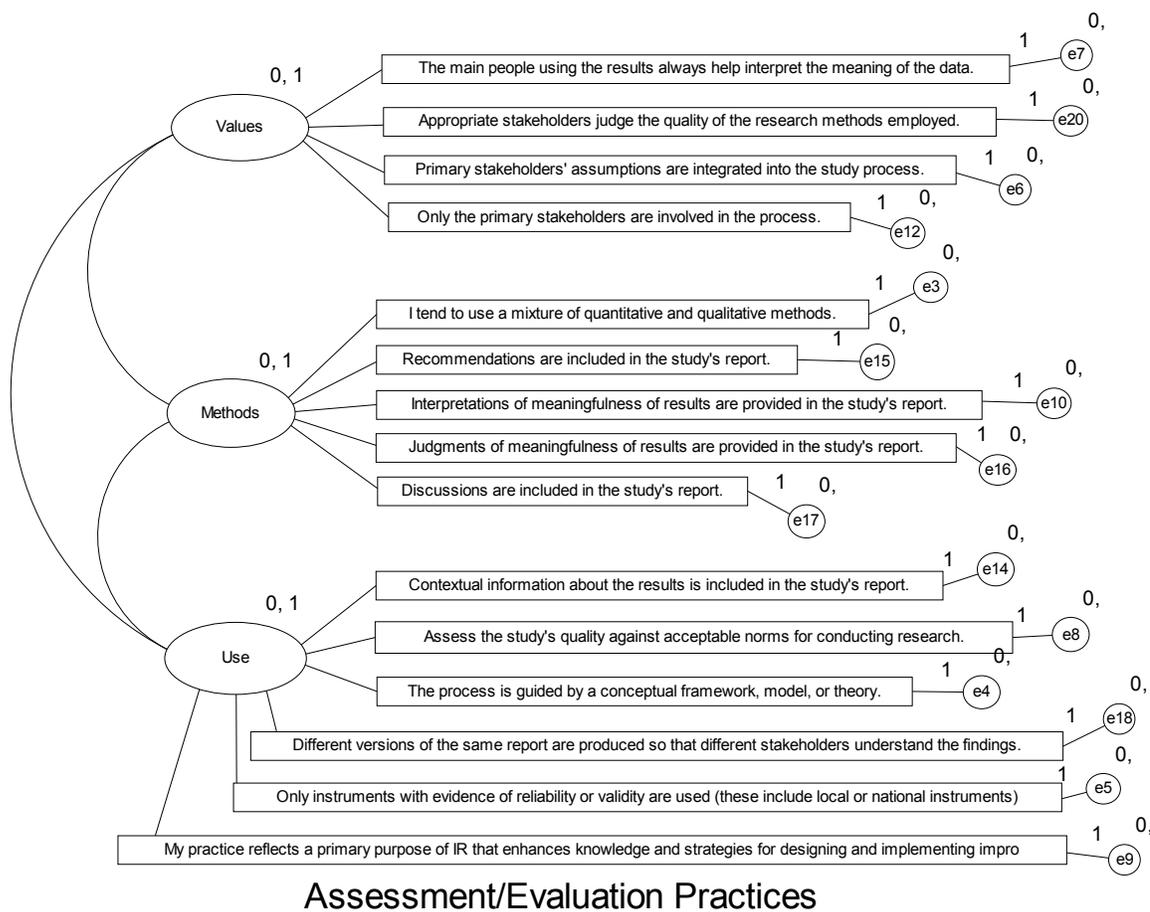
	3-Branch Model		
	Estimate	Standard Estimate	Standard Error
MainPeopleInterpret	0.469 <sup>***</sup>	.507	.046
JudgeQuality	0.592 <sup>***</sup>	.639	.046
PrimaryStakeAssumpt	0.486 <sup>***</sup>	.575	.042
Recommendations	0.735 <sup>***</sup>	.720	.040
Interpretations	0.731 <sup>***</sup>	.774	.036
Discussions	0.717 <sup>***</sup>	.719	.039
MultiVersions	0.355 <sup>***</sup>	.327	.050
IRenhancesKnowledge	0.577 <sup>***</sup>	.591	.042
ProvideContext	0.528 <sup>***</sup>	.651	.034
Reliability	0.329 <sup>***</sup>	.370	.040
Quant&Qual Methods	0.311 <sup>***</sup>	.292	.048
OnlyPrimaryStake	0.122 <sup>***</sup>	.144	.044
Judgments	0.793 <sup>***</sup>	.804	.037
Theory	0.455 <sup>***</sup>	.476	.042
AssessQual	0.637 <sup>***</sup>	.596	.046

<sup>\*\*\*</sup>Significant at  $p < 0.001$

**Table 14**  
**Covariance among Branches (Factors)**

3-Branch Model			Covariance	Standard Error
Values	↔	Use	.735***	.047
Methods	↔	Use	.845***	.028
Methods	↔	Values	.443***	.052

\*\*\* Significant at  $p < 0.001$



**Figure 2. Path Diagram for the 3-Branch Model (Assessment/Evaluation Practices)**

The 11 Assessment/Evaluation Practices items were explicitly constrained to load on only one branch in multiple branch models. In the 3-branch model, three items loaded only on Values, three items loaded only on Methods, and five items loaded only on Uses; in the 2-branch model, eight items loaded only on Methods plus Uses, and three items loaded only on Values; in the 1-branch model all items loaded on the Methods plus Uses plus Values. Results for each model tested are listed in Table 15.

**Table 15**

***Approaches to Conducting IR Model Results***

	$\chi^2$	df	p-value	CFI	RMSEA
3-Branch Model (Values, Methods, Uses)	113.40	41	< .001	.90	.08
2-Branch Model (Values, Methods + Uses)	198.63	43	< .001	.78	.11
1-Branch Model (Values + Methods + Uses)	350.33	44	< .001	.57	.16

As indicated in Table 15, the 3-branch model was the only model tested that met the fit criterion of CFI greater than or equal to 0.90 and RMSEA less than or equal to 0.08. Although this 2-branch model does not approach the fit criteria, a  $\chi^2$  difference test is conducted. The results indicated that the 3-branch model was statistically better fitting ( $\Delta\chi^2 = 85.23$ ,  $\Delta df = 2$ ,  $p < 0.01$ ).

In Table 16, the estimates, standardized estimates, and standard errors for each item in the 3-branch model are displayed and in Table 17 the covariance among branches (factors) are displayed.

**Table 16**

***Estimates, Standardized Estimates, and Standard Errors for Each Item in the 3-Branch Model***

	3-Branch Model		
	Estimate	Standard Estimate	Standard Error
StakeCollect	0.491 <sup>***</sup>	.493	.062
StakeAnalysis	0.829 <sup>***</sup>	.816	.060
StakeInterpret	0.756 <sup>***</sup>	.791	.057
Judgments	0.736 <sup>***</sup>	.747	.062
Interpretations	0.787 <sup>***</sup>	.828	.062
BuildsKnowledge	0.482 <sup>***</sup>	.608	.050
ReplicateStudies	0.326 <sup>***</sup>	.309	.071
Context	0.468 <sup>***</sup>	.452	.067
FocusDMs	0.339 <sup>***</sup>	.386	.058
CreatesChanges	0.490 <sup>***</sup>	.531	.059
NextSteps	0.566 <sup>***</sup>	.677	.053

<sup>\*\*\*</sup> Significant at  $p < 0.001$

As indicated in Table 17, factor covariance's (that can be interpreted as correlations) between Methods and Use, and Values and Use are statistically significant at the 0.001 level and are considered moderate in magnitude. Factor

covariance between Values and Methods was not statistically significant. Overall, the association among factors is low enough to provide some evidence of discriminate validity. In Figure 3 the path diagram for the 3-branch model is provided. Standardized regression weights, error variances, and intercepts are indicated on the path diagram.

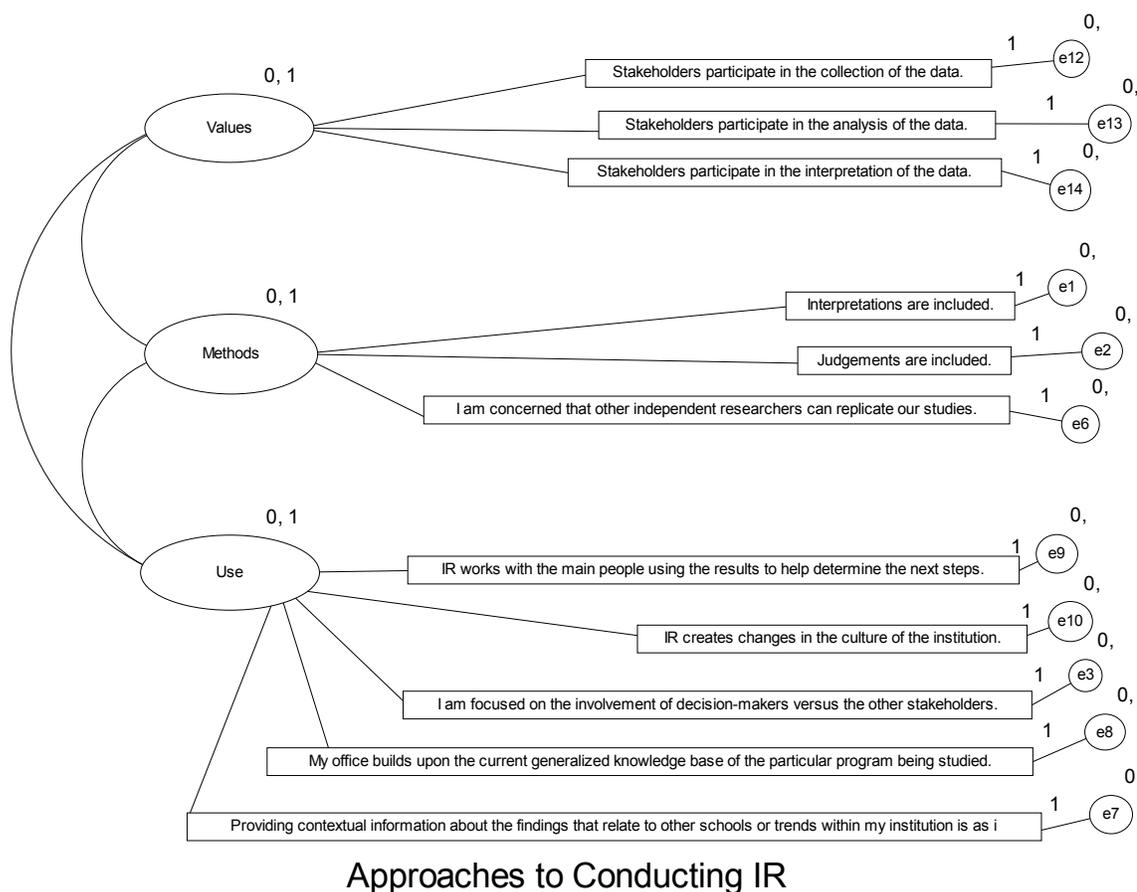
**Table 17**

***Covariance among Branches (Factors)***

3-Branch Model			Covariance	Standard Error
Methods	↔	Use	.581 <sup>***</sup>	.066
Values	↔	Methods	.050	.077
Values	↔	Use	.481 <sup>***</sup>	.068

<sup>\*\*\*</sup>Significant at  $p < 0.001$

Whereas each primary section of the survey was analyzed separately, associations in the scale scores from these different sections are expected. The correlation matrix among the different subscale scores is presented below (see Table 18). Examination of these correlations provides some indication of how the different areas relate to one another and how reasonable it is to assume that the sections of the survey should be treated as different areas where the evaluation theory tree structure emerges or whether a single evaluation theory tree structure applies to all three professional areas.



**Figure 3. Path Diagram for the 3-Branch Model (Approaches to Conducting IR)**

Shown in Table 18 are the Values and Use branches for each section of the survey. Note all are statistically significant and are moderately correlated (0.43 to 0.53 and 0.46 to 0.55 respectively). Furthermore, correlations for the Methods branches for each section of the survey were found to be statistically significant and have moderately high correlation. If corrections were made for reliability of the scales, the Methods branches would be nearly perfectly correlated. If larger sample sizes had been available, it may have been reasonable to test a single tree for all three professional areas.

**Table 18**

***Pearson Correlations***

		Values			Methods			Use		
		Core IR	Assess	Approach	Core IR	Assess	Approach	Core IR	Assess	Approach
Values	Core IR	<b>0.79</b>								
	Assess	.525**	<b>0.61</b>							
	Approach	.467**	.432**	<b>0.73</b>						
Methods	Core IR	.388**	.279**	.112	<b>0.85</b>					
	Assess	.320**	.325**	.189**	.759**	<b>0.78</b>				
	Approach	.288**	.224**	.126*	.723**	.690**	<b>0.63</b>			
Use	Core IR	.392**	.325**	.142*	.533**	.442**	.372**	<b>0.52</b>		
	Assess	.388**	.458**	.295**	.541**	.594**	.511**	.470**	<b>0.68</b>	
	Approach	.300**	.373**	.369**	.444**	.446**	.490**	.458**	.551**	<b>0.64</b>

\* significant at 0.05

\*\* significant at 0.01

Note: coefficient alpha's are indicated in bold along the diagonal

### MANOVA with Discriminant Follow Up Results

In Table 19 the mean branch scores by Evaluator and Non-Evaluator (IR) are displayed. Mean scores and standard deviations are indicated for each branch of each section of the survey. In all cases, participants classified as Evaluator had higher mean scores in all branches.

**Table 19**

#### *Mean Scores by Branch for Evaluator/IR*

		Eval-like IR	M	SD	N
Core IR Reporting	Values	IR	3.01	.86	250
		Eval-Like	3.46	.88	23
		Total	3.04	.87	273
	Methods	IR	3.13	.84	250
		Eval-Like	3.45	.77	23
		Total	3.16	.84	273
	Use	IR	3.67	.63	250
		Eval-Like	3.91	.61	23
		Total	3.69	.63	273
Assessment	Values	IR	3.08	.55	250
		Eval-Like	3.43	.68	23
		Total	3.11	.57	273
	Methods	IR	3.19	.78	250
		Eval-Like	3.65	.63	23
		Total	3.22	.77	273
	Use	IR	3.17	.59	250
		Eval-Like	3.63	.60	23
		Total	3.21	.61	273
Approaches	Values	IR	2.96	.81	250
		Eval-Like	3.21	.78	23
		Total	2.98	.80	273
	Methods	IR	2.88	.74	250
		Eval-Like	2.97	.79	23
		Total	2.88	.74	273
	Use	IR	3.41	.57	250
		Eval-Like	3.57	.55	23
		Total	3.42	.57	273

Box's M test for equal covariance across dependent variables is not significant,  $p = 0.540$  (see Table 20). Thus, the hypothesis that the covariances are equal across groups cannot be rejected.

**Table 20**

***Box's Test of Equality of Covariance Matrices***

Box's M	51.218
F	.964
df1	45
df2	4814.004
Sig.	.540

A one-way MANOVA (see Table 21) revealed a significant multivariate main effect for group membership in Eval/IR, Wilks'  $\lambda = .93$ ,  $F(9, 263.00) = 2.33$ ,  $p = .02$ , partial eta squared = .07. Power to detect the effect was quite high at .91. With the use of Wilks's criterion, the combined DVs (the nine branches) were significantly affected by group membership (Eval/IR). Thus the multivariate hypothesis that the mean on the composite was the same across groups was rejected.

Before evaluating the univariate tests on the branches, Levene's test for the assumption of equal variances for each of our dependent variables was considered. Table 22 provides the results for Levene's test. It is evident that the hypothesis of equal variances for each of our dependent variables cannot be rejected.

**Table 21****MANOVA Results for Eval/IR**

---

	<b>Effect</b>	<b>Value</b>	<b>F</b>	<b>Hypothesis df</b>	<b>Error df</b>	<b>Sig.</b>	<b>Partial Eta Squared</b>	<b>Noncent. Parameter</b>	<b>Observed Power</b>
	Pillai's Trace	.07	2.33	9.00	263.00	.02	.07	20.98	.91
Eval/IR	Wilks' Lambda	.93	2.33	9.00	263.00	.02	.07	20.98	.91
	Hotelling's Trace	.08	2.33	9.00	263.00	.02	.07	20.98	.91
	Roy's Largest Root	.08	2.33	9.00	263.00	.02	.07	20.98	.91

---

**Table 22*****Levene's Test of Equality of Error Variances***

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
Core_VALUES	.077	1	271	.781
Core_METHODS	.180	1	271	.672
Core_USE	.240	1	271	.625
Assess_VALUES	1.511	1	271	.220
Assess_METHODS	.717	1	271	.398
Assess_USE	.043	1	271	.836
Approach_VALUES	.008	1	271	.928
Approach_METHODS	.001	1	271	.981
Approach_USE	.578	1	271	.448

Table 23 provides abbreviate MANOVA results. Four branches were found to be statistically significant. All three mean branch scores for the Assessment/Evaluation Practices tree and the CORE IR Reporting Values branch were significantly different between participants classified as Evaluators and those classified as IR. Evaluator's mean score was found to be higher than IR on each branch of this tree; Evaluators mean score on the Assess\_Values branch was 3.43 versus 3.08 for IR; Evaluators mean score on the Assess\_Methods branch was 3.65 versus 3.19 for IR; Evaluators mean score on the Assess\_Use branch was 3.63 versus 3.17 for IR; Evaluators mean score on the Core\_Values branch was 3.46 versus 3.01 for IR. Although these findings were significant, the effect sizes were relatively small (partial eta-squared of .021 to .044).

**Table 23**

***Tests of Between-Subjects Effects***

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Eval vs IR	<b>Core_VALUES</b>	<b>4.338</b>	<b>1</b>	<b>4.338</b>	<b>5.844</b>	<b>.016</b>	<b>.021</b>	<b>5.844</b>	<b>.673</b>
	Core_METHODS	2.083	1	2.083	2.989	.085	.011	2.989	.406
	Core_USE	1.225	1	1.225	3.063	.081	.011	3.063	.415
	<b>Assess_VALUES</b>	<b>2.508</b>	<b>1</b>	<b>2.508</b>	<b>7.855</b>	<b>.005</b>	<b>.028</b>	<b>7.855</b>	<b>.798</b>
	<b>Assess_METHODS</b>	<b>4.509</b>	<b>1</b>	<b>4.509</b>	<b>7.706</b>	<b>.006</b>	<b>.028</b>	<b>7.706</b>	<b>.790</b>
	<b>Assess_USE</b>	<b>4.385</b>	<b>1</b>	<b>4.385</b>	<b>12.376</b>	<b>.001</b>	<b>.044</b>	<b>12.376</b>	<b>.939</b>
	Approach_VALUES	1.284	1	1.284	1.993	.159	.007	1.993	.290
	Approach_METHODS	.185	1	.185	.334	.564	.001	.334	.089
	Approach_USE	.561	1	.561	1.727	.190	.006	1.727	.258

### Discriminant Function Analysis

Discriminant function analysis was conducted using all nine branches to help describe differences between Evaluators and IR. Box's M and mean scores by group have already been presented. Four areas were found to make the largest contributions to group difference: Core Values and all three branches of the Assessment/Evaluation Theory Tree (Values, Methods, and Use). These same areas surfaced previously in the MANOVA analysis with similar results for F-values and significance and will not be repeated here. Because the dependent variable (Eval vs. IR) has only two possible categories, there is only one discriminate function. The canonical correlation in Table 24, indicates the percent of variation in the dependent discriminated by the independent variables. These nine branches explained approximately 0.272 of the variance in group membership. This amount of variation is considered low although, this finding is consistent with the modest effect sizes found in the univariate follow-up analysis.

**Table 24**

#### *Eigenvalues*

Function	Eigenvalue	Percent of Variance	Cumulative Percent	Canonical Correlation
1	.080	100.0	100.0	.272

This analysis is about describing differences rather than predicting group membership. As a result, the structure matrix is considered next (Table 25). By

identifying the largest absolute correlations associated with this discriminant function, descriptive differences between groups can be obtained. Table 25 indicates that the branches associated with the Assessment/Evaluation Practices tree tend to separate Evaluators from IR more than other branches and thus more than the other trees. However and more specifically, the Use branch of the Assessment and Evaluation Practices trees provides the most discriminating information (.757), followed next by the Values branch (.603), the Methods branch (.597), and by the Core IR Reporting Values branch (.520).

**Table 25**

***Structure Matrix***

	<b>Function</b>
	<b>1</b>
Assess_USE	.757
Assess_VALUES	.603
Assess_METHODS	.597
Core_VALUES	.520
Core_USE	.376
Core_METHODS	.372
Approach_VALUES	.304
Approach_USE	.283
Approach_METHODS	.124

### Multiple Linear Regression Results

The results from three multiple linear regression tests are presented below (Tables 26, 27, and 28). Table 26 presents the results of regressing the Core IR Reporting mean score on the Values branch items, in Table 27 the results of regressing the Core IR Reporting mean score on the Methods branch items are shown, and in Table 28 the results of regressing the Core IR Reporting mean score on the Use branch items are displayed. Statistically significant items are bolded and boxed. In all cases, the Variance Inflation Factor (VIF) is within acceptable ranges (less than 10).

Listed in Table 26 are the three Approaches items that were significantly related to the Core IR Reporting Values branch mean score: (a) stakeholders participating in the analysis, (b) interpretation of data, and (c) interpretations are included. Stakeholder items reside on the Values branch and interpretations being included reside on the Methods branch of the Approaches tree. The standardized coefficient for stakeholders participating in the analysis is nearly twice the size (0.259) of that of stakeholders interpret data and interpretations are included (0.147 and 0.144 respectively). Thus involving stakeholders in the analysis has a much greater impact on the Core IR Reporting Values branch.

The three Approaches items are significantly related to the Core IR Reporting Method branch mean score: (a) interpretations are included, (b) judgments are included, and (c) IR creates changes in the culture are shown in Table 27.

**Table 26**

***Mean CORE IR Reporting Values Branch: Approaches to IR Items Coefficients***

		Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics		
	Model	B	Standard Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	.751	.290		2.586	.010		
	Stakeholders participate in the collection of the data.	.072	.051	.082	1.399	.163	.732	1.367
	<b>Stakeholders participate in the analysis of the data.</b>	<b>.222</b>	<b>.059</b>	<b>.259</b>	<b>3.763</b>	<b>.000</b>	<b>.529</b>	<b>1.892</b>
	<b>Stakeholders participate in the interpretation of the data.</b>	<b>.134</b>	<b>.063</b>	<b>.147</b>	<b>2.132</b>	<b>.034</b>	<b>.526</b>	<b>1.900</b>
	<b>Interpretations are included.</b>	<b>.132</b>	<b>.063</b>	<b>.144</b>	<b>2.107</b>	<b>.036</b>	<b>.536</b>	<b>1.867</b>
	Judgments are included.	.095	.060	.107	1.585	.114	.553	1.809
	I am concerned that other independent researchers can replicate our studies.	.045	.045	.054	1.002	.317	.847	1.181
	IR works with the main people using the results to help determine the next steps.	.107	.065	.103	1.648	.101	.644	1.553
	IR creates changes in the culture of the institution.	.062	.055	.066	1.123	.262	.729	1.371

Model	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Standard Error	Beta	t	Sig.	Tolerance	VIF
I am focused on the involvement of decision-makers versus the other stakeholders.	-.105	.054	-.105	-1.926	.055	.841	1.188
My office builds upon the current generalized knowledge base of the particular program being studied.	-.004	.067	-.004	-.060	.952	.678	1.475
Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.	-.014	.047	-.016	-.288	.773	.793	1.261

**Table 27**

**Mean CORE IR Reporting Methods Branch: Approaches to IR Items Coefficients**

	Model	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Standard Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	.453	.194		2.340	.020		
	Stakeholders participate in the collection of the data.	.039	.034	.046	1.127	.261	.732	1.367
	Stakeholders participate in the analysis of the data.	.003	.039	.004	.076	.940	.529	1.892
	Stakeholders participate in the interpretation of the data.	-.057	.042	-.065	-1.351	.178	.526	1.900
	<b>Interpretations are included.</b>	<b>.362</b>	<b>.042</b>	<b>.411</b>	<b>8.635</b>	<b>.000</b>	<b>.536</b>	<b>1.867</b>
	<b>Judgments are included.</b>	<b>.345</b>	<b>.040</b>	<b>.406</b>	<b>8.661</b>	<b>.000</b>	<b>.553</b>	<b>1.809</b>
	I am concerned that other independent researchers can replicate our studies.	.000	.030	.001	.016	.987	.847	1.181
	IR works with the main people using the results to help determine the next steps.	.028	.043	.028	.654	.514	.644	1.553
	<b>IR creates changes in the culture of the institution.</b>	<b>.115</b>	<b>.037</b>	<b>.127</b>	<b>3.122</b>	<b>.002</b>	<b>.729</b>	<b>1.371</b>

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Standard Error	Beta			Tolerance	VIF
I am focused on the involvement of decision-makers versus the other stakeholders.	-.048	.036	-.050	-1.317	.189	.841	1.188
My office builds upon the current generalized knowledge base of the particular program being studied.	.056	.045	.053	1.258	.209	.678	1.475
Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.	.011	.032	.013	.336	.737	.793	1.261

Interpretation and judgments being included items reside on the Methods branch and IR creates changes in the culture reside on the Use branch of the Approaches tree. Including interpretations and judgments standardized coefficients are nearly equal (0.411 and 0.406, respectively) and are over three times larger than the coefficient for the IR creating changes item (0.127). Thus, these two items have greater impact on the Core IR Reporting Methods branch than the creating changes item.

Shown in Table 28 are the four Approaches items that were significantly related to the Core IR Reporting Use branch mean score: (a) IR works with the main people to determine the next steps, (b) IR creates changes in the culture, (c) focusing on decision-makers versus others, and (d) providing contextual information. All of these items reside on the Use branch of the Approaches tree. Of these, the standardize coefficient for the IR creates changes in the culture item is nearly twice the size as the others. Its standardized coefficient was 0.298. Thus, this item has a greater impact on the Core IR Reporting Use branch than all of the other items. It should be noted that the focusing on decision-makers versus others has a negative coefficient (-0.129) indicating that this item had a negative impact on Core IR Reporting Use branch mean score.

**Table 28**

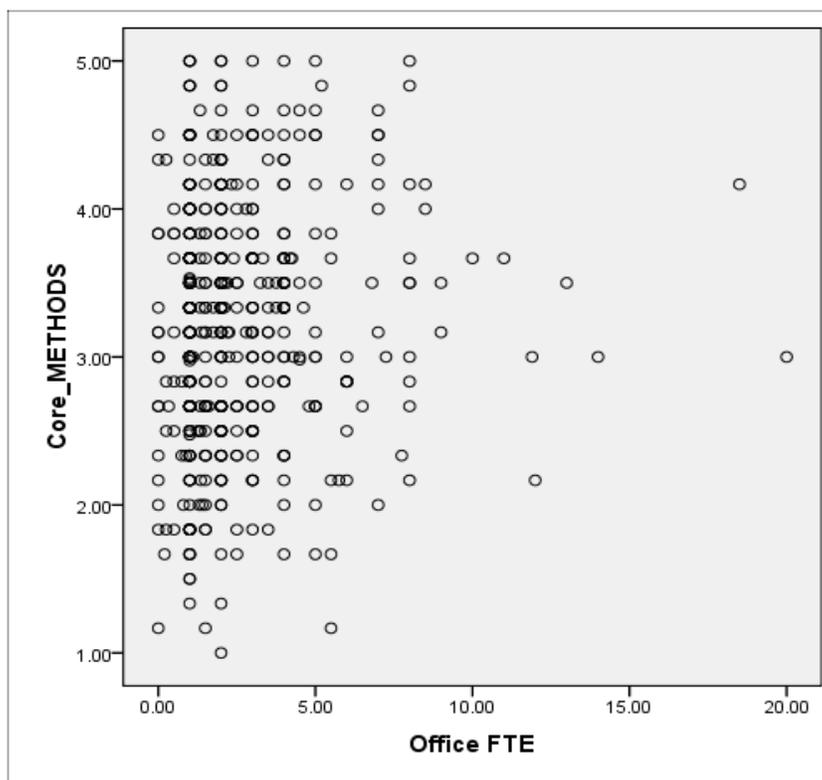
***Mean CORE IR Reporting Use Branch: Approaches to IR Items Coefficients***

		Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics		
Model		B	Standard Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	1.930	.211		9.149	.000		
	Stakeholders participate in the collection of the data.	-.043	.037	-.067	-1.156	.249	.732	1.367
	Stakeholders participate in the analysis of the data.	-.028	.043	-.044	-.641	.522	.529	1.892
	Stakeholders participate in the interpretation of the data.	.030	.046	.045	.657	.512	.526	1.900
	Interpretations are included.	.087	.046	.129	1.906	.058	.536	1.867
	Judgments are included.	.082	.043	.125	1.880	.061	.553	1.809
	I am concerned that other independent researchers can replicate our studies.	-.002	.033	-.004	-.069	.945	.847	1.181
	<b>IR works with the main people using the results to help determine the next steps.</b>	<b>.119</b>	<b>.047</b>	<b>.154</b>	<b>2.505</b>	<b>.013</b>	<b>.644</b>	<b>1.553</b>
	<b>IR creates changes in the culture of the institution.</b>	<b>.207</b>	<b>.040</b>	<b>.298</b>	<b>5.147</b>	<b>.000</b>	<b>.729</b>	<b>1.371</b>

Model	Unstandardized Coefficients		Standardized Coefficients		Collinearity Statistics		
	B	Standard Error	Beta	t	Sig.		
<b>I am focused on the involvement of decision-makers versus the other stakeholders.</b>	<b>-.094</b>	<b>.039</b>	<b>-.129</b>	<b>-2.389</b>	<b>.018</b>	<b>.841</b>	<b>1.188</b>
My office builds upon the current generalized knowledge base of the particular program being studied.	.057	.049	.071	1.181	.239	.678	1.475
<b>Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.</b>	<b>.100</b>	<b>.034</b>	<b>.161</b>	<b>2.911</b>	<b>.004</b>	<b>.793</b>	<b>1.261</b>

### Branch Scores: Function of FTE

Scatter plots were produced for FTE and each of the nine branch scores. No detectable trend was apparent. The scatter plot presented in Figure 4 below, FTE plotted with the CORE IR Reporting Methods branch, is representative of all scatter plot results. All scatter plot results, including the one below, are included in Appendix C.



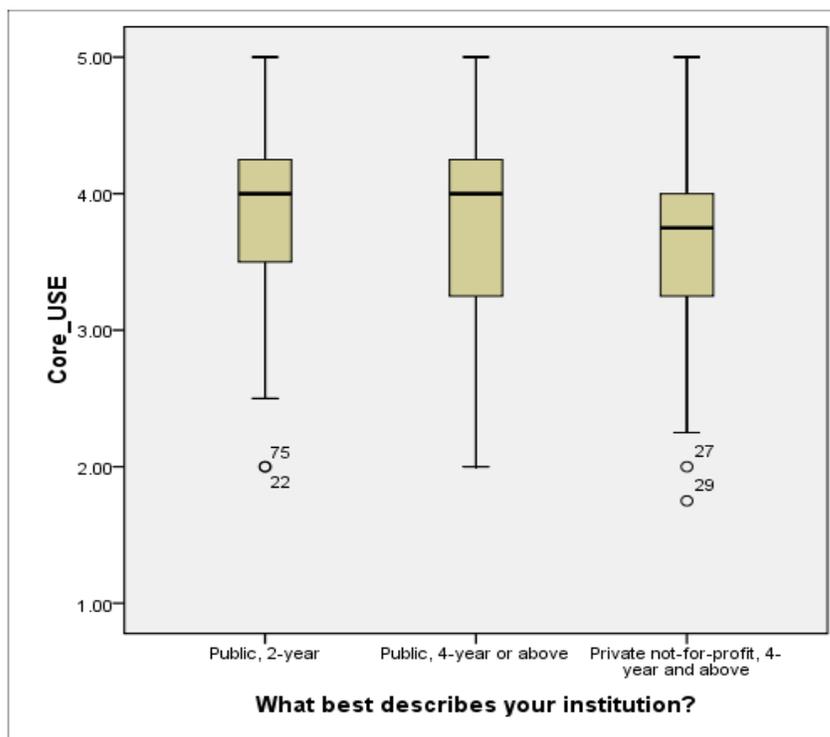
**Figure 4. Representative Scatterplot: FTE x CORE IR Reporting Methods Branch**

### **Branch Scores: Function of School Type**

Box plots were produced for the three Institutional types (non-profit Public 2-year, non-profit Public 4-year, and non-profit Private 4-year) and the nine mean branch scores. These three institutional types represent 543 of the 568 participants who chose an institutional type. Only the Core IR Reporting Use branch box plot revealed possible significant differences of branch placement based upon institutional type. Its box plot is presented below (see Figure 5) and is included in Appendix D with all other box plot results. Although the box plot does contain “outliers,” these are correct scores and were retained. The Private 4-year schools’ mean (3.63) was lower than either of the 2-year or 4-year Public schools’ mean (3.80 and 3.78 respectively). As a result of these findings, a one-way ANOVA was conducted to test for significant differences among institutional types.

Descriptive information is provided in Table 29. Private 4-year schools have a lower mean Core IR Reporting Use branch score (3.628) than Public 2-year or 4-year schools (3.803 and 3.798 respectively). For all practical purposes, the Public schools’ means were the same (approximately 3.80). Levene’s test for homogeneity of variances (see Table 30) was non-significant. As a result, the hypothesis that there are equal variances across school type was not rejected.

As shown in the one-way ANOVA table (see Table 31), there are statistically significant differences among the Core IR Reporting Use branch mean scores and college type ( $p = 0.008$ ).



**Figure 5. CORE IR Reporting Use Branch Box Plot**

**Table 29**

***Descriptive Information***

Core_USE						
					95% Confidence Interval for Mean	
	<i>N</i>	<i>M</i>	<i>SD</i>	Std. Error	Lower Bound	Upper Bound
Public, 2-year	169	3.803	.635	.049	3.706	3.899
Public, 4-year or above	140	3.798	.659	.056	3.688	3.908
Private not-for-profit, 4-year and above	234	3.628	.624	.041	3.547	3.708
Total	543	3.726	.641	.028	3.672	3.780

**Table 30*****Test of Homogeneity of Variances***

<b>Core_USE</b>			
<b>Levene Statistic</b>	<b>df1</b>	<b>df2</b>	<b>Sig.</b>
.312	2	540	.732

**Table 31*****One-Way ANOVA Results***

<b>Core_USE</b>					
	<b>Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>
Between Groups	3.984	2	1.992	4.918	.008
Within Groups	218.732	540	.405		
Total	222.716	542			

A Bonferroni post-hoc was performed to help identify the source of the significant differences among college type (see Table 32). However and as indicated previously, Public 2-year and 4-year school means were nearly identical. As a result, Public schools were not expected to be significantly from each other; whereas significant differences from Private 4-year schools are expected. As indicated in Table 32, there were no significant differences among the Public schools and Core Use branch mean scores. Public 2-year and 4-year schools mean scores on the Core IR Reporting Use branch were statistically significantly higher than Private 4-year schools mean score.

Table 32

*Multiple Comparisons*

Core_USE Bonferroni						
(I) What best describes your institution?	(J) What best describes your institution?	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Public, 2-year	Public, 4-year or above	.004	.073	1.000	-.170	.179
	Private not-for-profit, 4-year and above	.175*	.064	.020	.021	.329
Public, 4-year or above	Public, 2-year	-.004	.073	1.000	-.179	.170
	Private not-for-profit, 4-year and above	.171*	.068	.037	.007	.334
Private not-for-profit, 4-year and above	Public, 2-year	-.175*	.064	.020	-.329	-.021
	Public, 4-year or above	-.171*	.068	.037	-.334	-.007

## CHAPTER V

### CONCLUSIONS AND DISCUSSION

The primary purpose of this study was to map the Evaluation Theory Tree to Institutional Research. The first question was, “Does an Evaluation Theory Tree, containing Values, Methods, and Uses branches fit Institutional Research (IR)?” Confirmatory factor analysis was used to test model fit. The Evaluation Theory Tree structure was tested on three trees: (a) Core IR Reporting, (b) Assessment/Evaluation Practices, and (c) Approaches to Conducting IR. In each case a 3-branch model met the fit criteria. Because additional nested models existed, one and two branch models were also tested. Where appropriate, a  $\chi^2$  difference test was conducted to determine the better fitting model. The  $\chi^2$  difference tests revealed that 3-branch models were better fitting models in all cases. The overall conclusion was that a 3-branch model of the Evaluation Theory Tree appeared to fit IR.

When comparing covariance among branches (factors), two of the CORE IR Reporting branches are considered moderate (Use-Values and Methods-Values) and one is considered moderately high (Use-Methods). The covariance among branches on the Assessment/ Evaluation Practices tree indicated that one is moderate (Methods-Values), one is moderately high (Values-Use) and one is high (Methods-Use). Variance among branches on the Approaches to

Conducting IR tree indicated that two are moderate (Methods-Use and Values-Use) and one is not statistically distinguishable from zero (Values-Methods). Moderately high and high correlations among factors could be an indication that constructs represented by the model were not sufficiently differentiated. Although the correlation matrix presented indicated a moderately high correlation among methods branches, the overall findings indicated that having three separate trees for testing model fit appeared to be a reasonable approach. As such, these findings could help to develop an IR theory that is grounded in or built upon the Evaluation Theory Tree. As stated earlier, a theory of IR (or set of theories) can provide guidance in the selection of methods and how information is gathered and interpreted. An IR theory might also be help to align more IR offices with respect to function and responsibilities.

The second hypothesis addressed branch placement differences among self-identified Evaluators and IR practitioners. The MANOVA and the discriminant function analysis results indicated that there were statistically significant differences between these two groups. Whereas Evaluators had higher mean scores on all the tree branches, the Assessment/Evaluation Practices tree and the CORE IR Reporting Values branch yielded statistically significant differences for these groups. The results from the Assessment/Evaluation Practices tree makes some intuitive sense because that tree addresses areas where Evaluators are trained differently by virtue of their degree or membership in the American Evaluation Association—an organization that is focused on research, application,

and curriculum based on evaluation standards and practices. The importance of the Use branch on that tree as the most distinguishing factor between Evaluators and IR practitioners was surprising. One might have hypothesized that the Values branch would have been more of a distinguishing characteristic. As it turned out, evaluators, by virtue of their training, might be more adept at employing assessment/evaluation techniques that lead to greater use. If so, then IR members who have not had this type of training can seek out professional development opportunities to gain greater knowledge and application of Use techniques. The importance of the CORE IR Reporting Values branch as another distinguishing factor underscores the importance of incorporating the Values branch concepts into reporting. It appears that involving stakeholders into various aspects of reporting is an important process. If this is the case, then perhaps basic reporting is not so “basic”. Some within IR might not be aware of its complexity and could benefit from training in evaluation.

The third question examined which Approaches to Conducting IR are significantly related to Core IR Reporting scores. The Core IR Reporting items could be viewed as best practices and thus mean branch scores become a proxy for best practices by branch. Regressing Mean Core IR Reporting branch scores on items associated with the Approaches to Conducting IR tree yielded interesting results. The findings from the Values branch analysis indicated that two items associated with the Values branch of the Approaches tree were significant (*stakeholders participate in the analysis; stakeholders participate in*

*the interpretation of the data*) and one item from the Methods branch of the Approaches tree was also significant (*interpretations are included*). Whereas the two Values items are clearly an infusion of stakeholder values, perhaps the Methods item of *including interpretations* is a way for researchers to infuse their own interpretations. We should not lose sight of the fact that getting *stakeholders to participate in the analysis of the data* had a much larger impact on the mean scores of the Values branch. This result sends a signal about finding ways to involve stakeholders in the data analysis. For example, in an early stage of the study, IR could establish clear expectations to stakeholders about their involvement in various aspects of the data analysis.

Findings from the Core IR Reporting Methods branch analysis indicated that two items associated with the Methods branch of the Approaches tree were significant (*including judgments; including interpretations*) and one item from the Use branch of the Approaches tree was also significant (*IR creates changes in the culture of the institution*). In looking at those findings, *including judgments and interpretations* have a much larger impact on the Core IR Reporting Methods branch score. Such findings underscore the importance of including judgments and interpretations into IR approaches. While including interpretations about statistical findings into studies is relatively common practice, including judgments is not an abundant practice. Evaluation practices and procedures might provide some guidance to IR in this area.

Findings from the mean Core IR Reporting Use branch score indicate that four items from the Use branch of the Approaches tree were significant (*IR works with the main people to determine the next steps; IR creates changes in the culture; I am focused on the involvement of decision makers; providing contextual information*). Two items seemed particularly noteworthy: *IR creates changes* and *I am focused on decision makers*. The *IR creates changes in the culture of the institution* item has the largest impact on this branch score and it had nearly twice the impact as any other item. *I am focused in the involvement of decision makers* item had a negative impact on mean score. This impact suggests that decision makers are not the only essential stakeholders for consideration—particularly in the Use of reports. Similar to the findings in the Values branch, approaches that consider stakeholders in IR practices have significant implications. Higher education is a collegial environment and one in which faculty and staff members participate on a regular basis. Participation comes in many forms. But in particular, participation comes from the exchange of ideas (i.e., a learning process). Perhaps IR should consider itself in a role of educator/student and offer sessions that teach about IR as well as gathering information about what stakeholders need/want.

The results from the last two questions (is there a relationship between office FTE and branch scores and is there a relationship between institution type and branch scores) provided confirmation of the uniqueness of IR. Even though there are similarities such as producing a Fact Book, completing the Common

Data Set, and federal reporting, differences among IR offices are broad. Some offices engage in assessment, national surveys, and special studies. Other offices do not have those functions or purposes. The office duties, responsibilities, and functions are institutionally driven. Institutional history and need shapes any administrative office and thus, IR offices become unique from one another. Whereas uniqueness has its benefits, it also creates challenges (i.e., comparing offices or identifying offices that are considered exemplars). The evaluation theory tree may offer some value here. If institution type does not always help IR practitioners find peers, perhaps classification along evaluation branches could.

### **Limitations of this Study**

There are number of limitations to this study. First, there are inherent limitations to surveys and self reported information. For examples, there may be a response/non-response bias, scale interpretations across participates may not be consistent, participants could respond in a socially desirable way, and email invitations could be blocked by spam filters. Second, the reliability measures (coefficient alpha) were low on some scales. This low reliability may have limited the statistical power for research questions 2 through 4. Third, the Evaluator/IR sample size difference was large (with IR: evaluator ratio being approximately 10-to-1). Although statistically significant differences were detected on some measures with the low sample size for evaluators, results might have been different had there been larger sample sizes for evaluators. Fourth, while the

literature review helped guide item branch placement, the final branch association is somewhat subjective. Branch placement for some items might be considered easy; whereas other item placements could be considered difficult and thus, argumentative. For examples, some might argue that benchmarks, study limitations, or model guiding theory/processes should not have been associated with Use branches, but rather with Methods branches.

### **Future Research**

Future research could include improving the survey used in this study. Additional items or edited items might lead to improved psychometric properties of subscales (i.e., increased reliability of branch scores). More substantive changes are also possible. For example, future research could consider a Methods branch that transcends the separate area trees. Visually, Methods would be the trunk of the tree with the Values and Use branches attached to specific area branches (such as Reporting, Assessment, and Approaches). It should be noted that AMOS provides modification indices (MI). MI analyses provide a range of possible model modifications that can improve fit. For example, item-factor association(s) and correlation among items could allow for a better fitting model. Employing MI could have benefits in terms of constructing a better fitting model or helping to formulate enhancements to theory and practice. Even though this study considered Office FTE and institutional type as possible functions of branch placement, other categories (such as school selectivity, years as a director, or type of highest degree) should be considered in future studies.

Such findings could have implications in moving IR towards a discipline. Finally, future studies may need to consider over sampling of evaluators within the IR profession for a more balanced representation in Evaluator/IR membership analyses.

### **Conclusions**

This study has demonstrated that there are similar structures to professional practice in Evaluation and in IR. This similarity was demonstrated through the application of an Evaluation Theory Tree to IR. Practitioners of IR should find relevance from these results. For examples, understanding Evaluation techniques and processes may lead to enhanced use of reports or greater involvement of stakeholders by IR practitioners. If so, evaluation practices and procedures would have practical benefits for IR. Because of the current uniqueness of IR offices, skill sets vary from office to office. As such, IR professionals who expect to work at more than one institution during their career should consider developing skill sets that are transferable. Evaluation practice may be one aspect of that transferable skill set. Practitioners of evaluation should also find useful information in this study—such as the relationship between these professions and how each can benefit from studying the other. Clearly, more research into IR is needed to better understand how and why its processes and procedures can be improved.

Evaluation processes and procedures appear to have benefits that are applicable to IR. The exact benefits will depend upon the IR practitioner and

where they perceive their strengths and weaknesses reside. For example, is IR involving stakeholders sufficiently into the data analysis and interpretation of data? If not, then finding ways to enhance stakeholder participation becomes important. Self reflection for IR might also provide some clarity of functions. If believing that IR creates changes in the culture of an institution has an impact on the ways IR is conducted, and then finding ways to enhance and improve that belief, might lead to better ways of doing IR.

## REFERENCES

- AAUP. (2006, October 26). *Statement on Government of Colleges and Universities* (10<sup>th</sup> ed.). Retrieved March 23, 2008, from <http://www.aaup.org/AAUP/pubsres/policydocs/contents/governancestatement.htm>.
- Allen, M., & McGee, P. (2004). Measurement and evaluation in corporate universities. *New Directions for Institutional Research*, 124, 81-92.
- Alkin, M. C. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.
- American Evaluation Association. (2008a). *About us*. Retrieved July 24, 2008, from <http://www.eval.org/aboutus/organization/aboutus.asp>. Fairhaven, MA: Author.
- American Evaluation Association. (2008b). *University programs*. Retrieved July 24, 2008, from [http://www.eval.org/Training/university\\_programs.asp](http://www.eval.org/Training/university_programs.asp). Fairhaven, MA: Author.
- Arbuckle, J. L. (2006). Amos (Version 7.0) [Computer Program]. Chicago: SPSS.
- Association for Institutional Research. (2008). *AIR constitution and bylaws*. Retrieved July 25, 2008, from <http://www.airweb.org/?page=159>. Tallahassee, FL: Author.

- Bacharach, S. B., & Lawler, E. J. (1980). *Power and politics in organizations*. San Francisco: Jossey Bass.
- Bagshaw, M. (1999). Teaching institutional research to the learning-inhibited institution. *New Directions for Institutional Research, 104*, 73-82.
- Baker, G. A., & Roberts, C. L. (1989). Institutional effectiveness practices in community colleges in the southern region. *Community College Journal of Research and Practice, 13*(1), 45-61.
- Bledsoe, K. L., & Graham, J. A. (2005). The use of multiple evaluation approaches in program evaluation. *American Journal of Evaluation, 26*(3), 302-319.
- Bickman, L. (1994). An optimistic view of evaluation. *American Journal of Evaluation, 15*(3), 255-259.
- Birnbaum, R. (1988). *How colleges work: The cybernetics of academic organization and leadership* (1st ed.). San Francisco: Jossey-Bass.
- Borden, V. M. H., & Delaney, E. L. (1989). Information support for group decision making. *New Directions for Institutional Research, 64*, 49-59.
- Borland, K. W. (2001). Qualitative and quantitative research: A complementary balance. *New Directions for Institutional Research, 112*, 5-13.
- Brandon, P. R., Lindberg, M. A., & Wang, Z. (1993). Involving program beneficiaries in the early stages of evaluation: Issues of consequential validity and influence. *Educational Evaluation and Policy Analysis, 15*, 420-428.

- Brandon, P. R., Newton, B. J., & Harman, J. W. (1993). Enhancing validity through beneficiaries' equitable involvement in identifying and prioritizing homeless children's educational problems. *Evaluation and Program Planning, 16*, 287-293.
- Brandon, P. R. (1998). Stakeholder participation for the purpose of helping ensure validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation, 19*, 325-337.
- Brown, M. C. (2003). Emics and etics of researching black colleges: Applying facts and avoiding fallacies. *New Directions for Institutional Research, 118*, 27-40.
- Brown, J. A. (2008). Institutional research and external reporting: The major data seekers. *New Directions for Institutional Research, 141*, 87-96.
- Bryk, A. (Ed.). (1983). *Stakeholder-based evaluation, new directions for program evaluation* (Vol. 17). San Francisco: Jossey-Bass.
- Chase J. S. (1979). Time present and time past. *New Directions for Institutional Research, 23*, 15-22.
- Chase J. S., & Tetlow, W. L. (1979). Experience is the child of thought. *New Directions for Institutional Research, 23*, 23-31.
- Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation*. Thousand Oaks, CA: Sage.

- Chen, H. (1996). A comprehensive typology for program evaluation. *American Journal of Evaluation*, 17(2), 121-130.
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7-35.
- Christie, C. A. (2007). Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation*, 28(1), 8-25.
- Christie, C. A., & Azzam, T. (2005). What theorists say they do: A brief description of theorists' approaches. *New Directions for Evaluation*, 106, 15-26.
- Cummings, O. W., Nowakowski, J. R., Schwandt, T. A., Eichelberger, R. T., Kliet, K. C., Larson, C. L., & Knott, T. D. (1988). Business perspectives on internal/external evaluation. *New Directions for Evaluation*, 39, 59-74.
- Cooksy, L. J., & Caracelli, V. J. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Evaluation*, 26(1), 31-42.
- Cowley, W. H. (1959). *Two and a half centuries of institutional research*. In R. G. Axt & H. T. Sprague (Eds.), *College self-study: Lectures on institutional research*. Boulder, CO: Western Interstate Commission for Higher Education.

- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*, 397-418.
- Cousins, J., Donohue, J., & Bloom, G. (1996). Collaborative evaluation in North America: Evaluators' self-reported opinions, practices, and consequences, *Evaluation Practice, 17*(3), 207-226.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record, 64*, 672-683.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Cummings, O. W., Nowakowski, J. R., Schwandt, T. A., Eichelberger, R. T., Kliet, K. C., Larson, C. L., & Knott, T. D. (1988). Business perspectives on internal/external evaluation. *New Directions for Evaluation, 39*, 59-74.
- Datta, L. (2007). Looking at the evidence: what variations in practice might indicate. *New Directions for Evaluation, 113*, 35-54.
- Delbecq, A. L., & VandeVen, A. H. (1971). A group process model for problem identification and program planning. *Journal of Applied Behavioral Science, 7*(4), 466-491.
- Delaney, A. M. (1997). The role of institutional research in higher education: Enabling researchers to meet new challenges. *Research in Higher Education, 38*(1), 1-16.

- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Consideration for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-10. Available: <http://pareonline.net/pdf/v14n20.pdf>.
- Doi, J. I. (1979). The beginnings of a profession: A retrospective view. *New Directions for Institutional Research, 23*, 33-41.
- Donmoyer, R. (1990). Curriculum evaluation and the negotiation of meaning. *Language Arts, 67*, 274-286.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York: Taylor and Francis Group.
- Donaldson, S. I., & Christie, C. A. (2006). Emerging career opportunities in the transdiscipline of evaluation science. In S. I. Donaldson, D. E. Berger, & K. Pezdek (Eds.), *Applied psychology: New frontiers and rewarding careers*. Mahwah, NJ: Erlbaum.
- Dressel, P. L. (1981). The shaping of institutional research and planning. *Research in Higher Education, 14*(3), 229-258.
- Ehrenberg, E. H. (2005). Why universities need institutional researchers and institutional researchers need faculty members more than both realize. *Research in Higher Education, 46*(3), 349-363.
- Ewell, P. (1989). Enhancing information use in decision making. *New Directions for Institutional Research, 64*, 7-19.

- Fetterman, D. M. (1996). *Empowerment evaluation: An introduction to theory and practice*. In D. M. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment and accountability* (pp. 3-49). Thousand Oaks, CA: Sage.
- Fetterman, D. (2000). Summary of the STEP evaluation. *American Journal of Evaluation, 21*, 239–259.
- Fetterman, D. (2001). The Transformation of evaluation into a collaboration: A vision of evaluation in the 21<sup>st</sup> century. *American Journal of Evaluation, 22*, 381-385.
- Fincher, C. (1981). Is there an institution in institutional research? *Research in Higher Education, 14*(2), 179-182.
- Fincher, C. (1985). The art and science of institutional research. *New Directions for Institutional Research, 46*, 17-37.
- Fitzpatrick, J. L. (2004). Exemplars as case studies: Reflections on the links between theory, practice and context. *American Journal of Evaluation, 25*(4), 541-559.
- Fitzpatrick, J. L., Christie, C. A., & Mark, M. M., 2009. *Evaluation in action: Interviews with expert evaluators*. Thousand Oaks, CA: Sage.
- Forest, J. (2003). Service academies of the United States: Issues of context, curriculum and assessment. *New Directions for Institutional Research, 118*, 79-103.

- Gill, J. I., & Saunders, L. (1992). Conducting policy analysis in higher education. *New Directions for Institutional Research*, 76, 15-27.
- Greene, J. C. (2005a). A value-engaged approach for evaluating the Bunche-Da Vinci Learning Academy. *New Directions for Evaluation*, 106, 27-45.
- Greene, J. C. (2005b). Context. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 82-84). Thousand Oaks, CA: Sage.
- Green, B. L., Mulvey, L., Fisher, H. A., & Woratschek, F. (1996). Integrating program and evaluation values: A family support approach to program evaluation. *Evaluation Practice*, 17, 261-272.
- Greene, J. C., Lipsey, M. W., Schwandt, T. A., Smith, N. L. & Tharp, R. G. (2007). Methods choices are contextual, contingent, and political. *New Directions for Evaluation*, 113, 111-127.
- Grob, G. F. (2003). A truly useful bat is found in the hands of a slugger. *American Journal of Evaluation*, 24(4), 449-505.
- Henry, G. T. (2000). Why not use? *New Directions for Evaluation*, 88, 85-98.
- Henry, G. T. (2005). *In pursuit of social betterment: A proposal to evaluate the Da Vinci Learning Model*. *New Directions for Evaluation*, 106, 47-63.
- Henry, G. T., & Julnes, G. (1998). Values and realist evaluation. *New Directions for Evaluation*, 78, 53-72.
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and action. *American Journal of Evaluation*, 24(3), 293-314.

- House, E.R., 1986. Internal Evaluation. *American Journal of Evaluation*, v7, n1, pp. 63-64.
- Howard, R. D., & Borland, K. W. (2001). Integrating qualitative and quantitative information for effective institutional research. *New Directions for Institutional Research*, 112, 100-115.
- Hoyle, R. H. (1995). *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage.
- Impara, J., Plake, B., & Fager, J. (1993). Teachers' assessment background and attitudes toward testing. *Theory into Practice*, 32(2) 113- 117.
- Johnston, W. P. (1998). Increasing evaluation use: Some observations based on the results of the U.S. GAO. *New Directions for Evaluation*, 39, 75-84.
- Johnson, R. L., Willeke, M. J., & Steiner, D. J. (1998). Stakeholder collaboration in the design and implementation of a family literacy portfolio assessment. *American Journal of Evaluation*, Vol. 19, 339-353.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed). Thousand Oaks, CA: Sage.
- Julnes, G., & Rog, D. J. (2007). Current federal policies and controversies over methodology in evaluation. *New Directions for Evaluation*, 113, 1-12.
- Kendall-Tackett, K. (2005). Section 1: Conducting an evaluability assessment. Retrieved September 10, 2005, from <http://www.fourh.umn.edu/evaluation/family/Section1/Page-06.html>

- Knapp, M. (1995). How shall we study comprehensive, collaborative services for children and families? *Educational Researcher*, 24(4), 5-16.
- Kinser, K. (2003). Diversity within the virtual classroom. *New Directions for Institutional Research*, 118, 69-77.
- Kinnick, M. K. (1985). Increasing use of student outcomes information. *New Directions for Institutional Research*, 47, 93-109.
- Knight, W. E., Moore, M. E., & Coperthwaite, C. A. (1999). Knowledge, skills, and effectiveness in institutional research. *New Directions for Institutional Research*, 104, 31-39.
- Knott, T. D. (1988). The impact of major consumer groups on health care evaluation. *New Directions for Evaluation*, 39, 33-46.
- Lane, J. E. (2003). Studying community colleges and their students: Context and research issues. *New Directions for Institutional Research*, 118, 51-67.
- Lane, J. E., & Brown, M. C. (2004). The importance of acknowledging context in institutional research. *New Directions for Institutional Research*, 124, 93-103.
- Lane, J. E., & Brown, M. C. (2003). Looking backward to see ahead: Implications for research, policy, and practice. *New Directions for Institutional Research*, 118, 105-111.
- Lane, J. E., Brown, M. C., & Pearcey, M.-A. (2004). Transnational campuses: Obstacles and opportunities for institutional research in the global education market. *New Directions for Institutional Research*, 124, 49-62.

- Langdon, E. A., & Giovengo, A. L. (2003). Survey and assessment in women's colleges: An issue of context. *New Directions for Institutional Research*, 118, 7-26.
- Lawrenz, F., Gullickson, A., & Toal, S. (2007). Dissemination: Handmaiden to evaluation use. *American Journal of Evaluation*, 28(3), 275-289.
- Lindquist, S. B. (1999). A profile of institutional researchers from AIR national membership surveys. *New Directions for Institutional Research*, 104, 41-50.
- Lipsey, M. W. (2001). Re: Unsolved problems and unfinished business. *American Journal of Evaluation*, 22, 325-328.
- Love, A. J. (1991). *Internal evaluation: building organizations from within*, Sage Publications, Newbury Park, California.
- Love, A. J. (1983). The organizational context and the development of internal evaluation. *New Directions for Evaluation*, 20, 5-22.
- MacDonald, B. (1976). Evaluation and the control of education. In D. Tawney (Ed.), *Curriculum evaluation today: Trends and implications* (pp. 125-136). London: McMillian.
- Mangano, M. (1991). Using the media to disseminate evaluation results. *American Journal of Evaluation*, 12(2), 115-120.
- Mark, M. M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, 22(3), 457-479.

- Mark, M. M., Greene J. C., & Shaw, I. F. (Eds.). (2006). *The Sage handbook of evaluation*. Thousand Oaks, CA: Sage.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*, 35-56.
- Mark, M. M., Henry, G. T., & Julnes, G. (1999). Toward an integrative framework for evaluation practice. *American Journal of Evaluation, 20*(2), 177-198.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.
- Mathison, S. (2005). *Encyclopedia of evaluation*. Thousand Oaks: CA. Sage.
- McLaughlin, G. W., & McLaughlin, J. S. (1989). Barriers to information use: The organizational context. *New Directions for Institutional Research, 64*, 21-33.
- McLaughlin, G. W., McLaughlin, J. S., Muffo, J. A. (2001). Using qualitative and quantitative methods for complementary purposes: A case study. *New Directions for Institutional Research, 112*, 15-44.
- Mertens, D. M. (2001). Inclusivity and transformation: Evaluation in 2010. *American Journal of Evaluation, 22*(3), 367-374.
- Mohan, R., & Sullivan, K. (2006). Managing the politics of evaluation to achieve impact. *New Directions for Evaluation, 112*, 7-23.
- Mowbray, C. T. (1988). Getting the system to respond to evaluation findings. *New Directions for Evaluation, 39*, 47-58.

- Muffo, J. A. (1999). A comparison of findings from regional studies of institutional research offices. *New Directions for Institutional Research, 104*, 51-59.
- Muscatello, D. B. (1988). Developing and agenda that works: The right choice at the right time. *New Directions for Evaluation, 39*, 21-31.
- Nagel, S. S. (1983, October). *Factors facilitating the utilization of policy evaluation research*. Paper presented at the annual meeting of the Evaluation Research Society, Chicago, IL.
- National Center for Education Statistics. Retrieved March 1, 2011, from <http://nces.ed.gov> website:  
<http://nces.ed.gov/ipeds/glossary/index.asp?id=283>.
- Ortiz, A. M., & Boyer, P. (2003). Student assessment in tribal colleges. *New Directions for Institutional Research, 118*, 41-49.
- Patton, M. Q. (1987). *Utilization-focused evaluation* (2nd ed.) Newbury Park, CA: Sage.
- Patton, M. Q. (1988). Integrating evaluation into a program for increased utility and cost-effectiveness. *New Directions for Evaluation, 39*, 85-94.
- Patton, M. Q. (1996). A world larger than summative and formative. *Evaluation Practice, 17*(2), 131-144.
- Patton, M. Q. (1997). Toward distinguishing empowerment evaluation and placing it in a larger context: Evaluation practice. *American Journal of Evaluation, 18*(2), 147-163.

- Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22, 329-336.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks: CA. Sage.
- Patton, M. Q. (2007). Process use as a usefulism. *New Directions for Evaluation*, 116, 99-112.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Pawson, R., & Tilley, N. (2001). Realistic evaluation bloodlines. *American Journal of Evaluation*, 22, 317–324.
- Peterson, M. W. (1985). Institutional research: An evolutionary perspective. *New Directions for Institutional Research*, 46, 5-15.
- Peterson, M. W. (1999). The role of institutional research: From improvement to redesign. *New Directions for Institutional Research*, 104, 83-103.
- Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. *American Journal of Evaluation*, 18, 209-225.
- Ray, M. (2006). Choosing a truly external evaluator. *American Journal of Evaluation*; 27; 372-377.
- Raykov, T. & Marcoulides, G. A. (2006), *A First Course in Structural Equation Modeling* (2<sup>nd</sup> ed.). Mahwah: NJ. Lawrence Erlbaum Associates, Inc.

- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks: CA. Sage.
- Russ-Eft, D., & Presill, H. (2001). *Evaluation in organizations: A systematic approach to enhanced learning, performance, and change*. Cambridge, MA: Perseus.
- Sample, J. A. (1984, March). Nominal group technique: An alternative to brainstorming. *Journal of Extension*. Retrieved August 17, 2008, from <http://www.joe.org/joe/1984march/iw2.html>.
- Saupe, J. L. (1990). *The functions of institutional research* (2nd ed.) Tallahassee, FL: Association for Institutional Research.
- Saupe, J. L. (2005). *How old is institutional research and how did it develop?* Unpublished paper. Remarks at the Annual MidAIR conference, November 2005. Retrieved July 11, 2008, from <http://www.mid-air.org/images/How%20Old%20is%20Institutional%20Research%20and%20How%20Did%20It%20Develop.pdf>.
- Schalock, R. L. (2001). *Outcome-based evaluation* (2nd ed.). New York: Kluwer Academic/Plenum Publishers.
- Schein, E. H. (1985). *Organizational Culture and Leadership*. (1<sup>st</sup> ed.) San Francisco: Jossey-Bass.
- Schmidlein, F. A. (1999). Emerging perspectives on organizational behavior: Implications for institutional research. *New Directions for Institutional Research*, 104, 61-71.

- Scriven, M. (1991). *Evaluation thesaurus*. Thousand Oaks, CA: Sage.
- Scriven, M. (1996). Types of evaluation and types of evaluators. *American Journal of Evaluation*, 17(2), 151-161.
- Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary vision. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Vision for the new millennium* (pp. 19-42). Mahwah, NJ: Lawrence Erlbaum.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage.
- Shadish, W. R., & Luellen, J. K. (2005). History of evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 183-186). Thousand Oaks, CA: Sage.
- Smith, M. F. (1988). Evaluation utilization revisited. *New Directions for Evaluation*, 39, 7-19.
- Smith, N. C., & Jackson, J. F. L. (2004). Religious institutions in the United States: Research challenges. *New Directions for Institutional Research*, 124, 31-48.
- Smith, E. R., & Tyler, R. W. (1942). *Appraising and recording student progress*. New York: HarperCollins.
- Sonnichsen, R.C., 1987. An Internal Evaluator Responds to Ernest House's Views on Internal Evaluation. *American Journal of Evaluation*, v8, n4, pp. 34-36.

- Southern Association of Colleges and Schools. (2007). *Quality Enhancement Plan handbook*. Decatur, GA: Commission on Colleges.
- Sridhara, S. (2003). Introduction to special section on "What is a Useful Evaluation?" *American Journal of Evaluation*, 24(4), 483-487.
- Stake, R. E. (2001). A problematic heading. *American Journal of Evaluation*, 22, 349-354.
- Stronach, I. (2001). From history observed one may prophesy. *New Directions for Evaluation*, 92, 59-72.
- Stufflebeam, D. L. (2000). Lessons in Contracting for Evaluations. *American Journal of Evaluation*, 21, 293-314.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). *Educational evaluation and decision-making*. Itasca, IL: F. E. Peacock.
- Stufflebeam, D. L. (1998). Conflicts between standards-based and postmodernist evaluations: Toward rapprochement. *Journal of Personnel Evaluation in Education*, 12(3), 287-296.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, & applications*. San Francisco, CA: Jossey-Bass.
- Sun, J. C. (2004). Professional schools: Research and assessment involving multiple constituencies. *New Directions for Institutional Research*, 124, 5-29.

- Tabachnick, B. G. & Idell, L. S. (2001). *Using Multivariate Statistics* (4<sup>th</sup> ed.), Needham Heights, MA: A Pearson Education, Inc.
- Taut, S. M., & Alkin, M. C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation*, 24(2), 213-226.
- Teodorescu, D. (2006). Institutional researchers as knowledge managers in universities: Envisioning new roles for their IR profession. *Tertiary Education and Management*, 12, 75-88.
- Terenzini, P. T. (1993). On the nature of institutional research and the knowledge of skills it requires. *Research in Higher Education*, 34, 1-10.
- Terenzini, P. T. (1999). On the nature of institutional research and the knowledge of skills it requires. *New Directions for Institutional Research*, 104, 21-29.
- Tetlow, W. L. (1979). From history observed one may prophesy. *New Directions for Institutional Research*, 23, 1-13.
- Tetlow, W. L. (1984). Using microcomputers for planning and management support. *New Directions for Institutional Research*, 44, 93-97.
- Thomas, C. L. (2000). Commentary: Politics, context, and integrity. *American Journal of Evaluation*, 21(2), 269-273.
- Tinto, V. (1988). *Stages of student departure: Reflections on the longitudinal character of student leaving*. *Journal of Higher Education*, 59, 438-455.

- Torres R. T., Preskill H. S., & Piontek, M. E. (1997). Communicating and reporting: Practices and concerns of internal and external evaluators. *American Journal of Evaluation* 1997; 18; 105-125.
- Trochim, W. M. (2004, November). *Randomized experiments: Lessons from the system of clinical trials in medical research*. Plenary presentation at the meeting of the American Evaluation Association, Atlanta, GA.
- Volkwein, J. F. (1999). *The four faces of institutional research*, *New Directions for Institutional Research*, 104, pp. 9-19.
- Weiss, C. H., Murphy-Graham, E., Petrosina, A., & Gandhi, A. G. (2008). The fairy godmother—and her warts: Making the dream of evidence-based policy come true. *American Journal of Evaluation*, 29, 343-347.
- Wholey, J. S. (2001). Managing for results: Roles for evaluators in a new management era. *American Journal of Evaluation*, 22, 29-47.
- WordNet® 3.0. Retrieved July 24, 2008, from Dictionary.com website: <http://dictionary.reference.com/browse/effectiveness>.
- Yang, H., & Shen, J. (2006). When is an external evaluator no longer external? *American Journal of Evaluation*, 27(3), 378-382.
- Zamani-Gallaher, E. M. (2004). Proprietary schools: Beyond the issues of profit. *New Directions for Institutional Research*, 124, 63-79.

## APPENDIX A

### SURVEY INSTRUMENT

Page 1

Office Reporting Practices

#### Instructions

The following statements are related to office reporting practices in institutional research. These reports are typically produced for institutional personnel (i.e, senior staff members, directors, department chairs) and can be routine or ad-hoc in nature. Please indicate how similar each statement is to the way your office actually conducts or produces reports.

Remember, we are interested in your actual practice rather than your ideal practice.

When conducting/producing reports...

{Choose one}

( ) None 1, ( ) Very little 2, ( ) Some 3, ( ) Quite a bit 4, ( ) Very much 5

- Methodological procedures are regularly included in reports.
- Benchmark or comparison figures are regularly included.
- Recommendations are regularly included in our reports.
- My office regularly includes the main people using the reports to help interpret the meaning of the data.
- Unless it is highly sensitive information, the reports are available for anyone to see.
- Stakeholders help to judge the appropriateness of the methods employed.
- Stakeholders help to judge the quality of the methods employed.
- My office regularly provides interpretations of the results.
- My office regularly provides judgments in reports.
- My office seeks feedback on our reports (i.e., the usefulness of or needed improvements to).
- Limitations are regularly included in our reports.
- Suggestions are regularly included in our reports.
- Discussions are regularly included in our reports.

Page 2

Approximately how many assessment or evaluation related activities does your office conduct during a typical academic year?

Assessment and evaluation activities can include special in-depth or on-going studies focused on programs, services, concepts, or constructs. Such studies typically involve measuring the results from rubrics, surveys, tests, focus groups, or interviews.

These activities can include assistance at various operational levels such as a unit, division, school, or institutional-level, and can include academic and non-academic programs and services.

{Choose one}

None,  1 to 4,  5 to 8,  9 to 12,  13 to 16,  17 to 20,  21 or more

Page 3

### Assessment/Evaluation Activities and Approaches

#### Instructions

The following statements are related to activities or approaches when conducting assessment/evaluation studies. Please indicate how similar each statement is to your actual assessment/evaluation practice.

Remember, we are interested in your actual practice rather than your ideal practice.

Primary stakeholders are persons who are directly involved in the delivery of the program or service (i.e., department chair-department faculty, director-subordinates)

When conducting assessment/evaluation related activities...

{Choose one}

( ) None 1, ( ) Very little 2, ( ) Some 3, ( ) Quite a bit 4, ( ) Very much 5

- My primary method of choice is quantitative.
- My primary method of choice is qualitative.
- I tend to use a mixture of quantitative and qualitative methods.
- The process is guided by a conceptual framework, model, or theory.
- Only instruments with evidence of reliability or validity are used (these include local or national instruments)
- Primary stakeholders' assumptions are integrated into the study process.
- The main people using the results always help interpret the meaning of the data.
- A primary role for my office is to assess the study's quality against acceptable norms for conducting research.
- My practice reflects a primary purpose of IR that enhances knowledge and strategies for designing and implementing improvements to programs and services, and not just to assess effectiveness.
- Interpretations of meaningfulness of results are provided in the study's report.

Page 4

- Continued - Assessment/Evaluation Activities and Approaches

Remember, we are interested in your actual assessment/evaluation practice rather than your ideal practice.

Decision-makers are stakeholders who have the authority to authorize changes to the program or service (i.e., Provost, Dean, department chair, director)

Primary stakeholders are persons who are directly involved in the delivery of the program or service (i.e., department chair-department faculty, director-subordinates)

Relevant stakeholders include decision-makers, primary stakeholders and customers/clients/students of the program or service

When conducting assessment/evaluation related activities...

{Choose one}

( ) None 1, ( ) Very little 2, ( ) Some 3, ( ) Quite a bit 4, ( ) Very much 5

- Only decision-makers are involved in the process.
- Only the primary stakeholders are involved in the process.
- All relevant stakeholders are involved in the process.
- Contextual information about the results is included in the study's report.
- Recommendations are included in the study's report.
- Judgments of meaningfulness of results are provided in the study's report.
- Discussions are included in the study's report.
- Different versions of the same report are produced so that different stakeholders understand the findings.
- Appropriate stakeholders judge the appropriateness of the research methods employed.
- Appropriate stakeholders judge the quality of the research methods employed.

Page 5

### Helping Stakeholders Conduct Assessment/Evaluation

#### Instructions

The following statements are related to helping stakeholders when conducting assessment/evaluation activities.

Remember, we are interested in your actual assessment/evaluation practice rather than your ideal practice.

My approach to conducting assessment/evaluation related activities is to act as a facilitator helping stakeholders...

{Choose one}

None 1,  Very little 2,  Some 3,  Quite a bit 4,  Very much 5

- Assess where the program is at the study's inception.
- Develop their own mission.
- Establish their own goals.
- Establish strategies for accomplishing their own goals.
- Establish evidence to document the strategies that will assist in reaching their goals.
- Measure how well they are doing in reaching their goals.

Page 6

Approximately how many measurement instruments (i.e., surveys, tests, rubrics) does your office help to create during an academic year?

{Choose one}

- None
- 1 to 4
- 5 to 8
- 9 to 12
- 13 to 16
- 17 to 20
- 21 or more

Page 7

### Practices in Developing & Testing Instruments

#### Instructions

The following statements are related to practices when developing and testing instruments. Please indicate how similar each statement is to your actual instrument development/testing practice.

Remember, we are interested in your actual practice rather than your ideal practice.

When developing an instrument ...

{Choose one}

None 1,  Very little 2,  Some 3,  Quite a bit 4,  Very much 5

- My office reviews research or literature findings about the topic at hand.
- My office conducts sound methodological pilot tests before going "live".
- My office conducts tests to see if the sample is representative of the population being studied.
- My office examines the correlation among variables that we think describe the construct(s) being measured.
- My office conducts reliability tests on the instruments we use.
- My office compare the results to some external criteria (i.e. student reported GPA versus actual GPA, religiousness questions with a group known to be highly religious).

Page 8

### Your Approach in Conducting IR

#### Instructions

The following statements are related to approaches when conducting IR. Please indicate how similar each statement is to your actual approach in practicing IR.

Remember, we are interested in your actual approach rather than your ideal approach.

Decision-makers are stakeholders who have the authority to authorize changes to the program or service (i.e., Provost, Dean, department chair, director)

Primary stakeholders are persons who are directly involved in the delivery of the program or service (i.e., department chair-department faculty, director-subordinates)

Relevant stakeholders include decision-makers, primary stakeholders and customers/clients/students of the program or service

In my approach to conducting institutional research (IR) ...

{Choose one}

( ) None 1, ( ) Very little 2, ( ) Some 3, ( ) Quite a bit 4, ( ) Very much 5

- Interpretations are included.
- Judgments are included.
- I am focused on the involvement of decision-makers versus the other stakeholders.
- It is more important to provide decision-makers what they need versus other stakeholders.
- I am concerned that the results from our studies are useful for other populations or other academic settings.
- I am concerned that other independent researchers can replicate our studies.
- Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.

Page 9

- Continued- Your Approach in Conducting IR

Remember, we are interested in your actual approach rather than your ideal approach.

In my approach to conducting institutional research (IR) ...

{Choose one}

None 1,  Very little 2,  Some 3,  Quite a bit 4,  Very much 5

- My office builds upon the current generalized knowledge base of the particular program being studied.
- IR works with the main people using the results to help determine the next steps.
- IR creates changes in the culture of the institution.
- Stakeholders are included in the study's process (i.e., collection, analysis, or interpretation of the data).
- Stakeholders participate in the collection of the data.
- Stakeholders participate in the analysis of the data.
- Stakeholders participate in the interpretation of the data.
- Conclusions contain a mixture of facts and values.

What is your level of agreement with the following statement:

The institutional researcher's interpretation of the findings and conclusions can be unbiased.

{Choose one}

Strongly Disagree 1,  Disagree 2,  Neither agree or disagree 3,  Agree 4,  Strongly Agree 5

Page 10

Demographic Information

Gender:

{Choose one}

Female

Male

Race/ethnicity:

Do you consider yourself to be Hispanic/Latino?

{Choose one}

Yes

No

In addition, select one or more of the following racial categories to describe yourself:

{Choose one}

American Indian or Alaska Native

Asian

Black or African American

Native Hawaiian or Pacific Islander

White

Age:

{Choose one}

Less than 30

30 to 35

36 to 40

41 to 45

46 to 50

51 to 55

56 or older

Highest degree obtained:

{Choose one}

High School

Some college

Bachelors

Masters

Doctorate

Your highest degree is in:

{Choose one}

Anthropology

Art

Biology

Business Administration

Computer Science

Economics

Education

English

Evaluation Research Methods

Higher Education

- History
- Human Services
- Information Science
- Mathematics
- Music
- Nursing/Medicine
- Political Science
- Public Administration
- Psychology
- Public Health
- Public Policy
- Religion
- School Administration
- Sociology
- Statistics
- Other [            ]

What best describes your relationship to IR?

{Choose one}

- IR is my full-time job
- IR is a part of my full-time job
- IR is a part-time job

How many years have you been conducting institutional research?

{Choose one}

- Less than 1
- 1 to 4
- 5 to 8
- 9 to 12
- 13 to 16
- 17 to 20
- 21 or more

How many years have you been a director (or higher level) of institutional research?

{Choose one}

- I am not the Director of IR
- Less than 1
- 1 to 4
- 5 to 8
- 9 to 12
- 13 to 16
- 17 to 20
- 21 or more

How would you rate your current IR knowledge and skills?

{Choose one}

- Very good
- Good
- Average
- Fair
- Poor

Do you teach college level courses in institutional research or courses that are related to institutional research?

{Choose one}

- Yes, but not in an IR certificate program
- Yes, I teach within an IR certificate program
- No

What best describes the selectivity of your institution?

(SAT = Math + Critical Reading)

{Choose one}

- Open: Generally open to all w/ H.S. diploma or GED
- Liberal: Average ACT 17-18, Average SAT 870-990, Majority admitted from bottom 50% H.S. Class
- Traditional: Average ACT 19-21, Average SAT 1000-1070, Majority admitted from top 50% H.S. Class
- Selective: Average ACT 22-26, Average SAT 1080-1220, Majority admitted from top 25% H.S. Class
- Highly Selective: Average ACT 27 +, Average SAT 1230 +, Majority admitted from top 10% H.S. Class

Including yourself, approximately how many fulltime equivalent (FTE) professionals are in the IR office? (please provide a single figure such as 1.33, 2.50, 4.00, 11.75)

{Enter text answer}

[ ]

Are you now or have you ever been a member of the American Evaluation Association?

{Choose one}

- Yes
- No

What best describes your institution?

{Choose one}

- Administrative Unit (i.e., a system or central coordinating office)
- Private for-profit, less-than 2-year
- Private for-profit, 2-year
- Private for-profit, 4-year or above
- Public, less-than 2-year
- Public, 2-year
- Public, 4-year or above
- Private not-for-profit, less-than 2-year
- Private not-for-profit, 2-year
- Private not-for-profit, 4-year and above
- Other [                    ]

Please use the space below to provide any additional comments.

{Enter answer in paragraph form}

[ ]

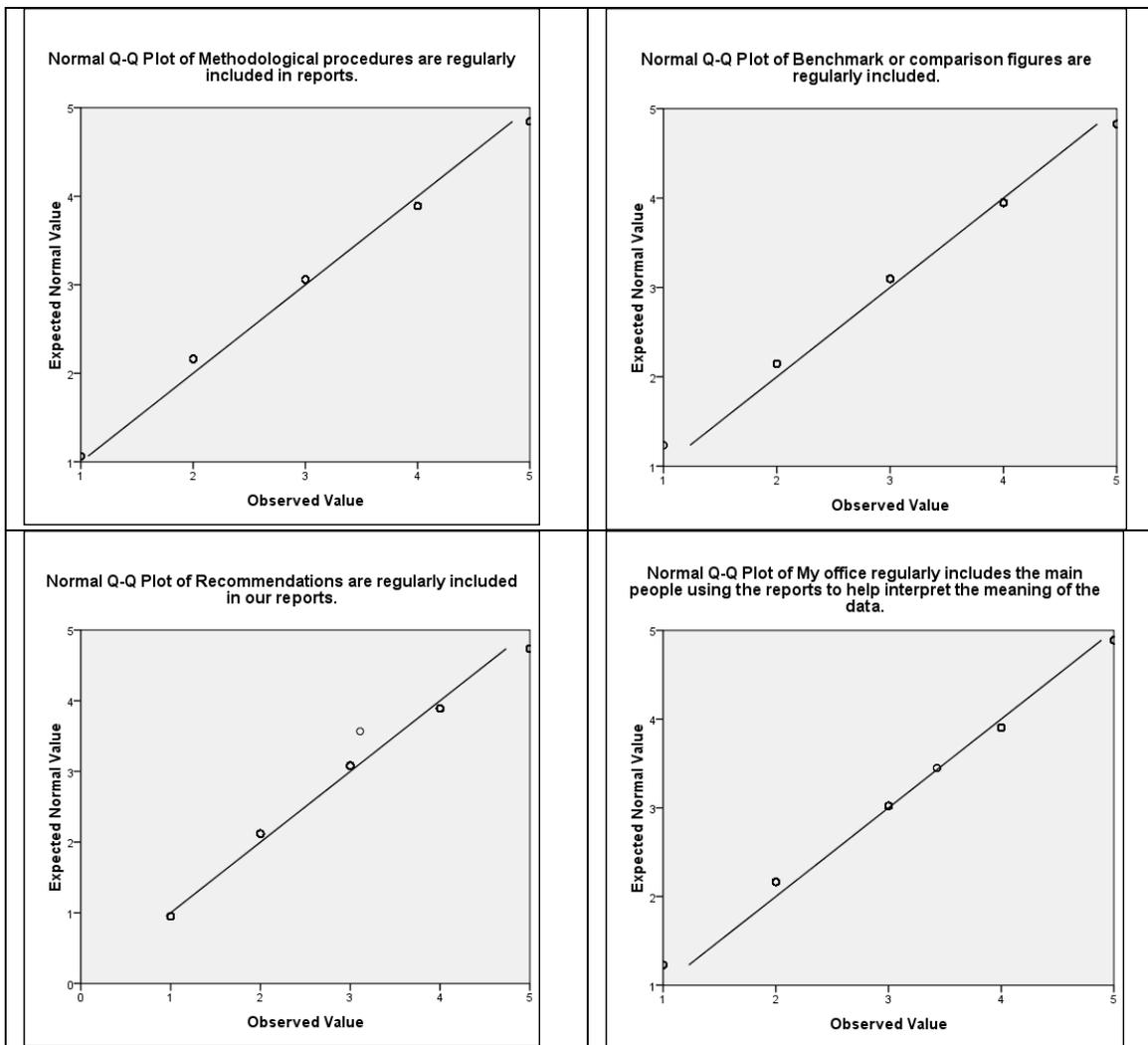
Thank you for your time.

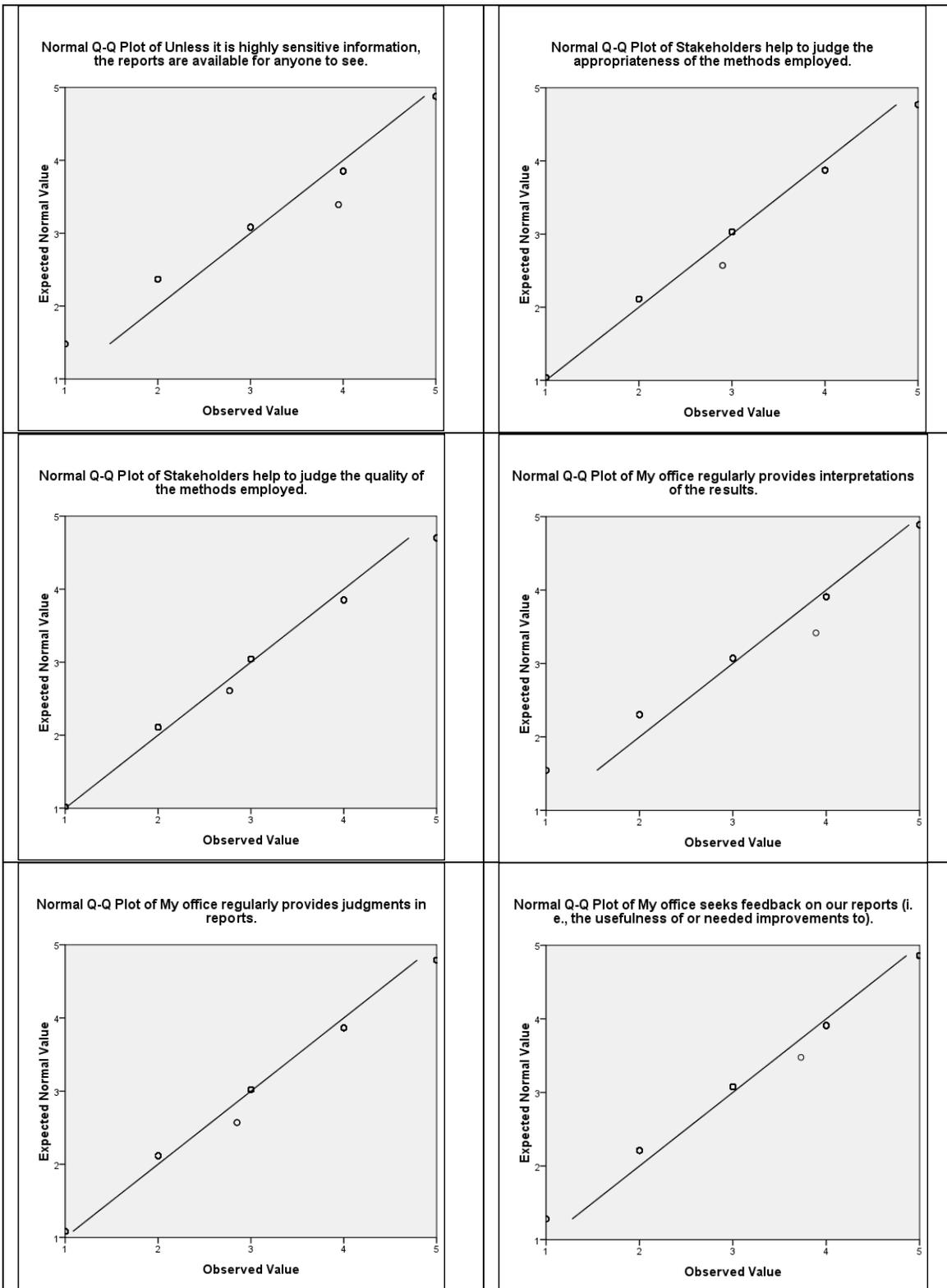
Please click the FINISH button to submit your responses.

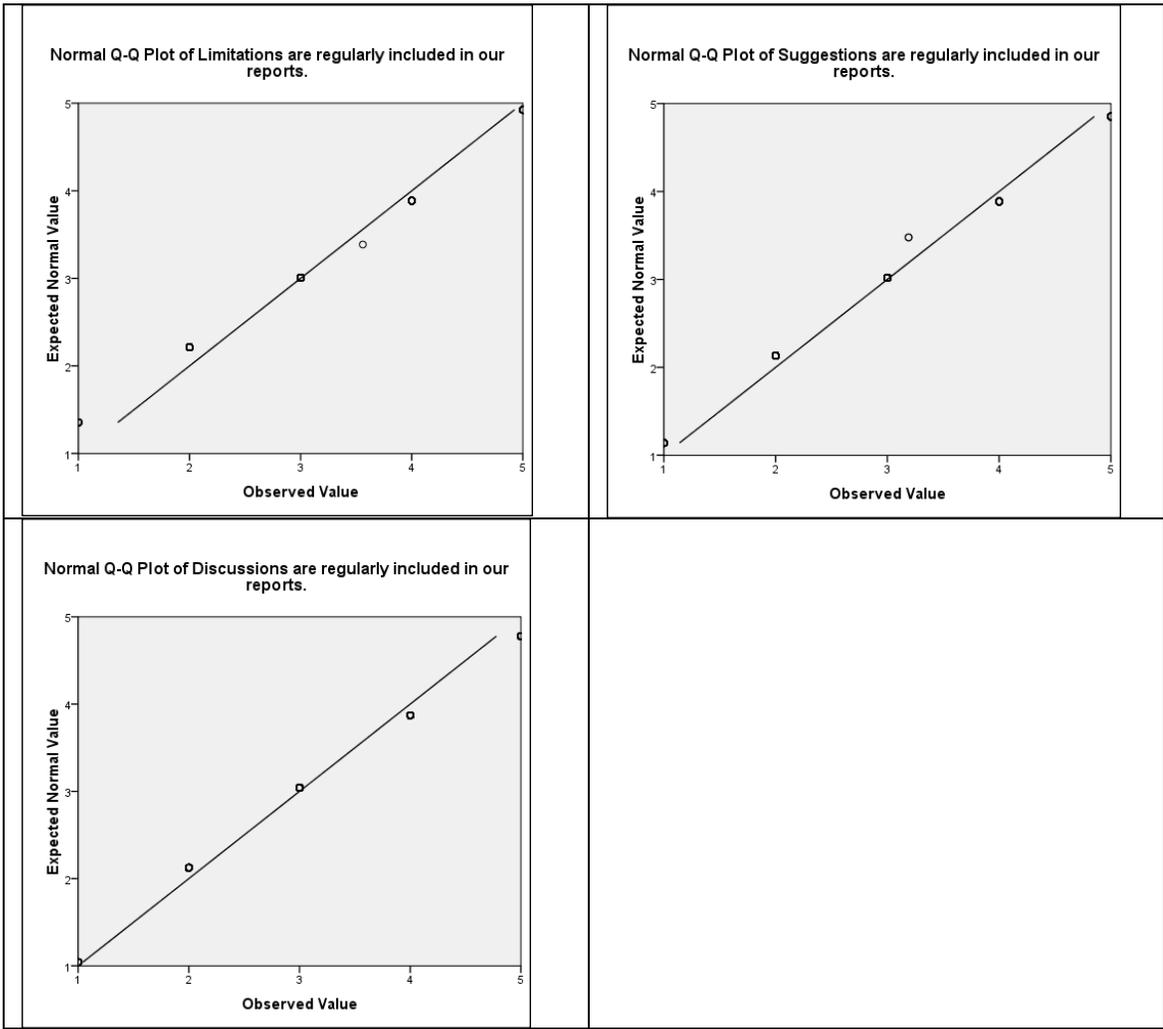
**APPENDIX B**

**Q-Q PLOTS**

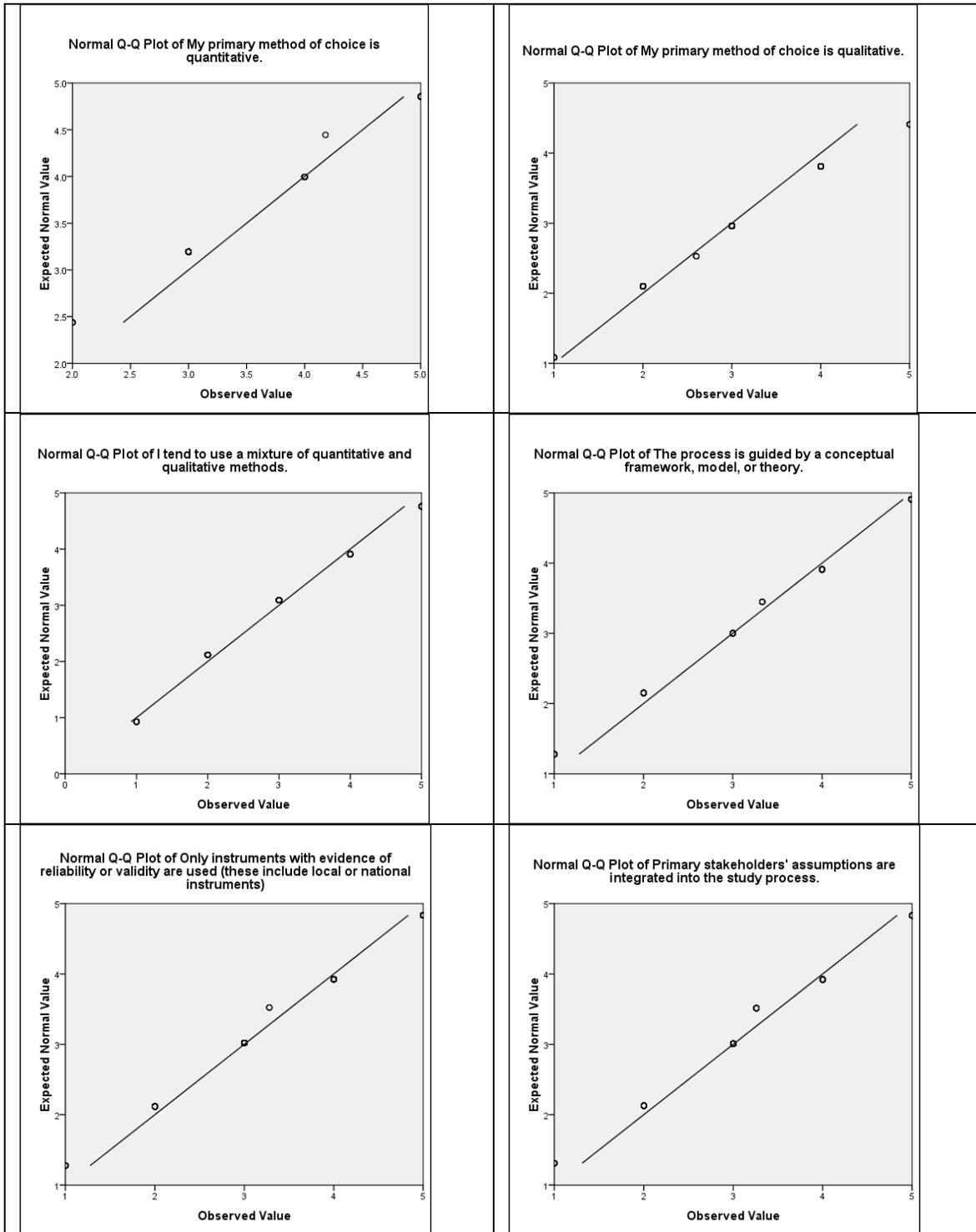
**CORE IR REPORTING**



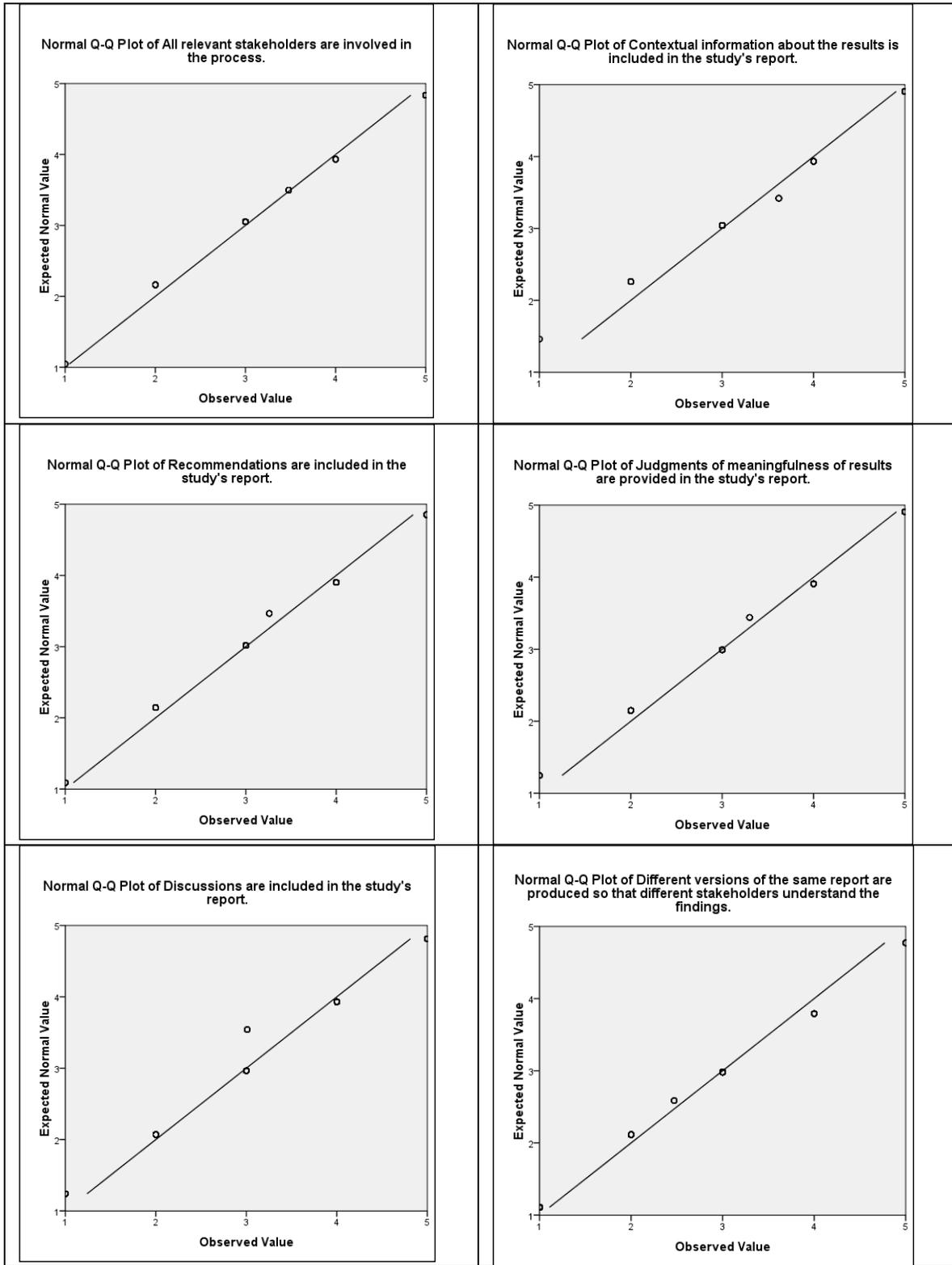


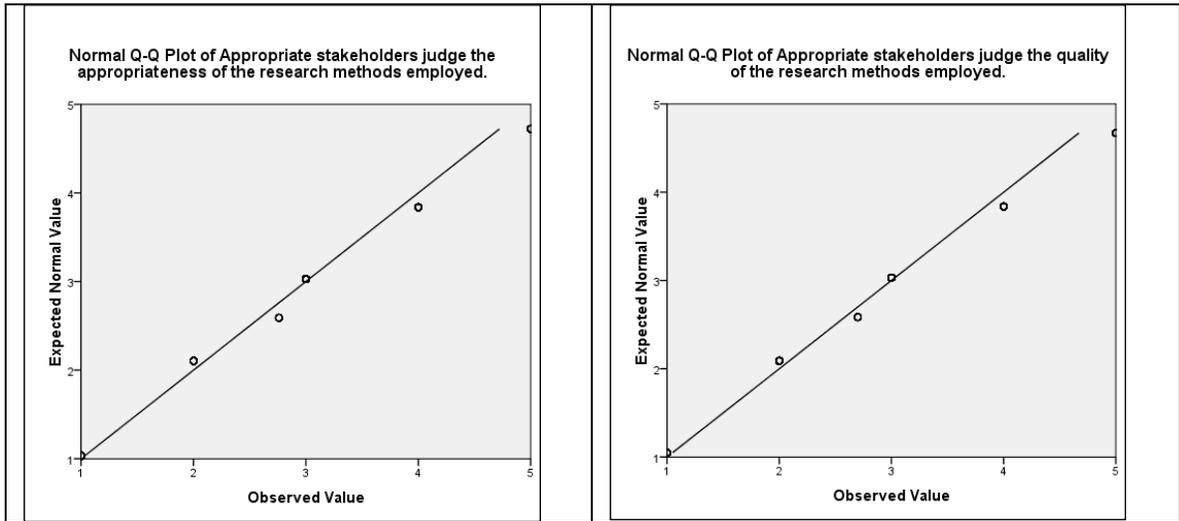


### ASSESSMENT/EVALUATION PRACTICES

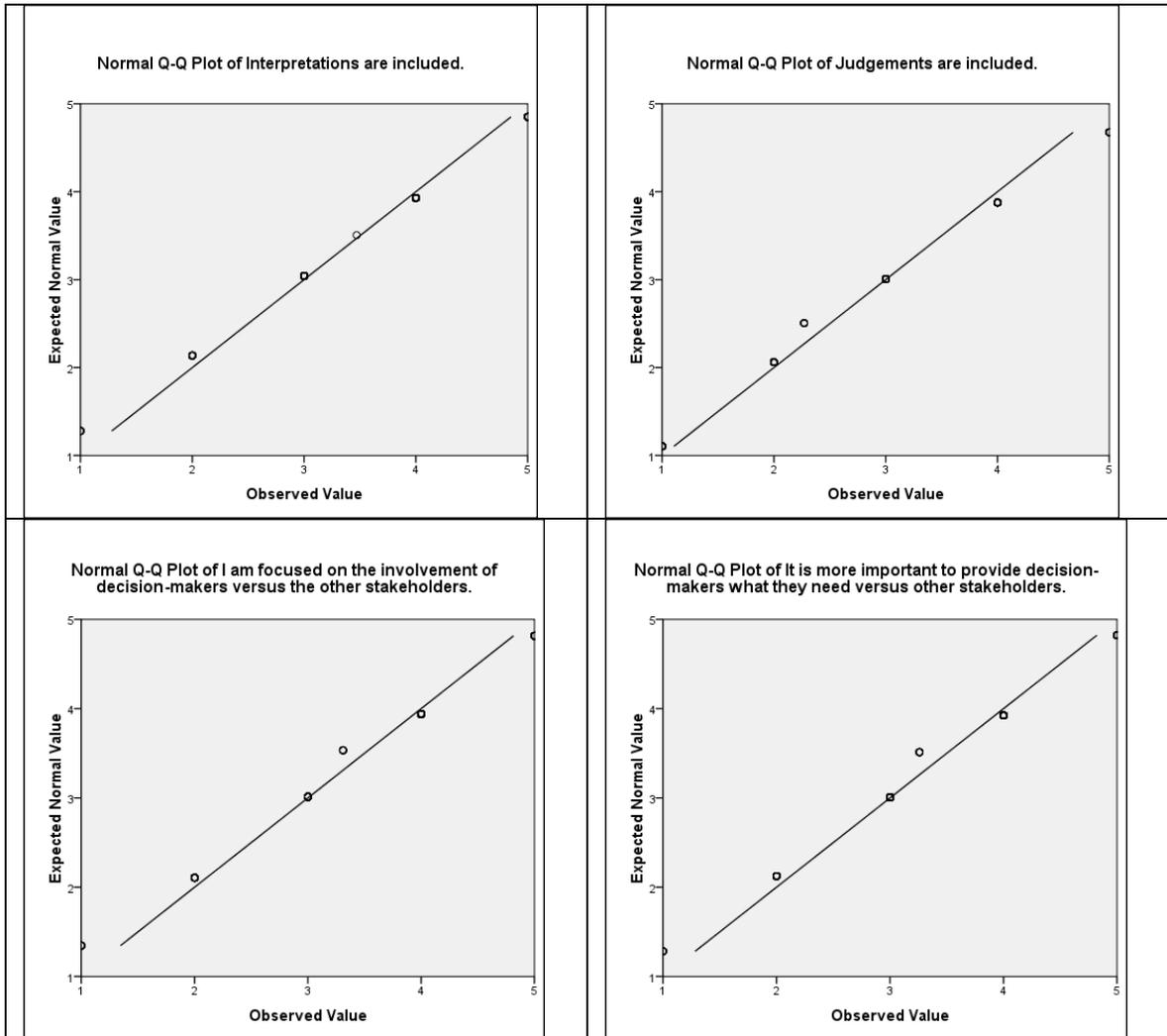


<p><b>Normal Q-Q Plot of The main people using the results always help interpret the meaning of the data.</b></p>		<p><b>Normal Q-Q Plot of A primary role for my office is to assess the study's quality against acceptable norms for conducting research.</b></p>	
<p><b>Normal Q-Q Plot of My practice reflects a primary purpose of IR that enhances knowledge and strategies for designing and implementing improvements to programs and services, and not just to assess effectiveness.</b></p>		<p><b>Normal Q-Q Plot of Interpretations of meaningfulness of results are provided in the study's report.</b></p>	
<p><b>Normal Q-Q Plot of Only decision-makers are involved in the process.</b></p>		<p><b>Normal Q-Q Plot of Only the primary stakeholders are involved in the process.</b></p>	

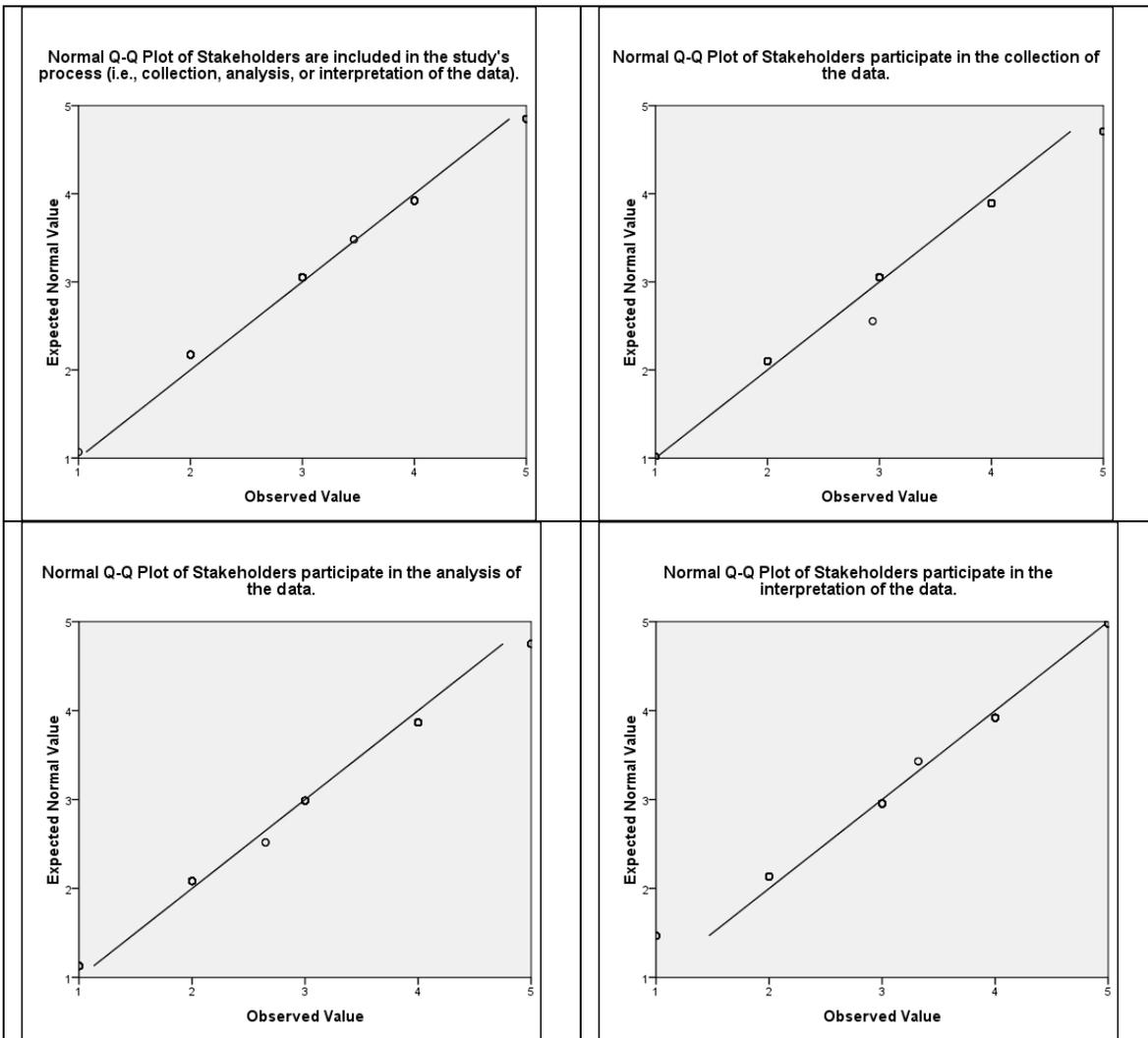




**APPROACHES TO CONDUCTING IR**



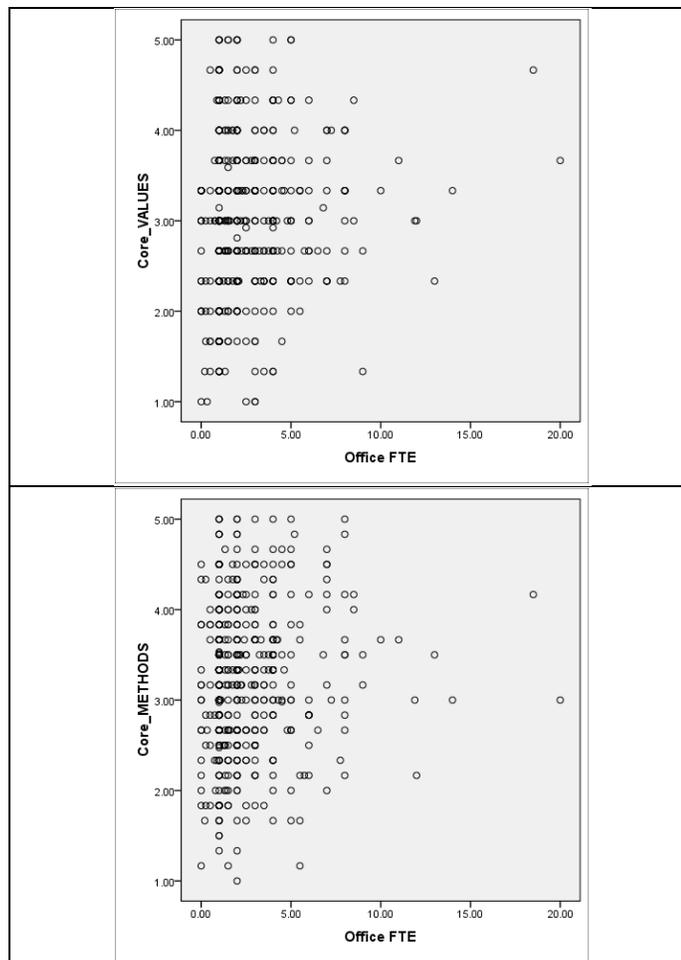
<p>Normal Q-Q Plot of I am concerned that the results from our studies are useful for other populations or other academic settings.</p>	<p>Normal Q-Q Plot of I am concerned that other independent researchers can replicate our studies.</p>
<p>Normal Q-Q Plot of Providing contextual information about the findings that relate to other schools or trends within my institution is as important as the findings themselves.</p>	<p>Normal Q-Q Plot of My office builds upon the current generalized knowledge base of the particular program being studied.</p>
<p>Normal Q-Q Plot of IR works with the main people using the results to help determine the next steps.</p>	<p>Normal Q-Q Plot of IR creates changes in the culture of the institution.</p>

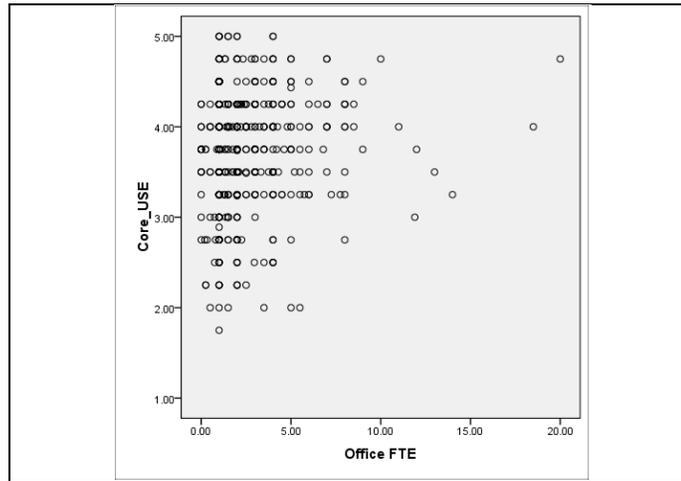


# APPENDIX C

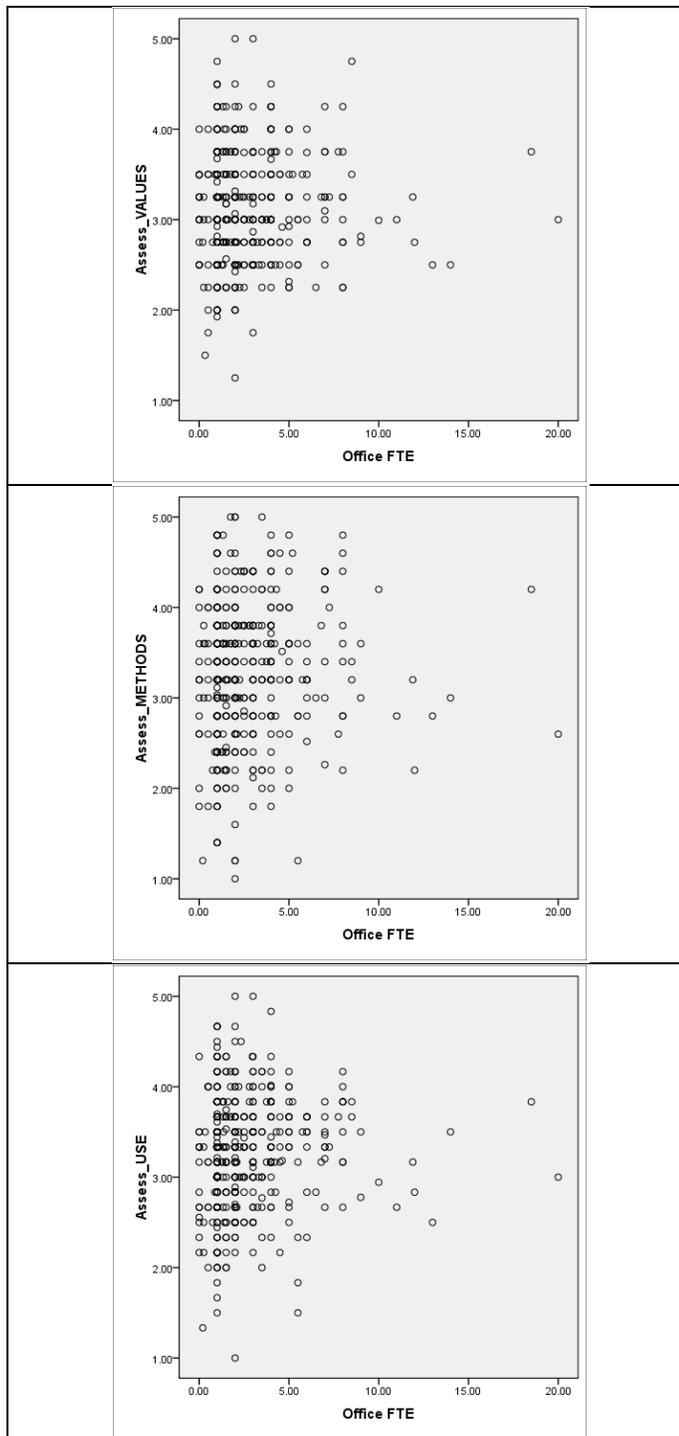
## SCATTER PLOTS

### CORE IR Reporting Tree

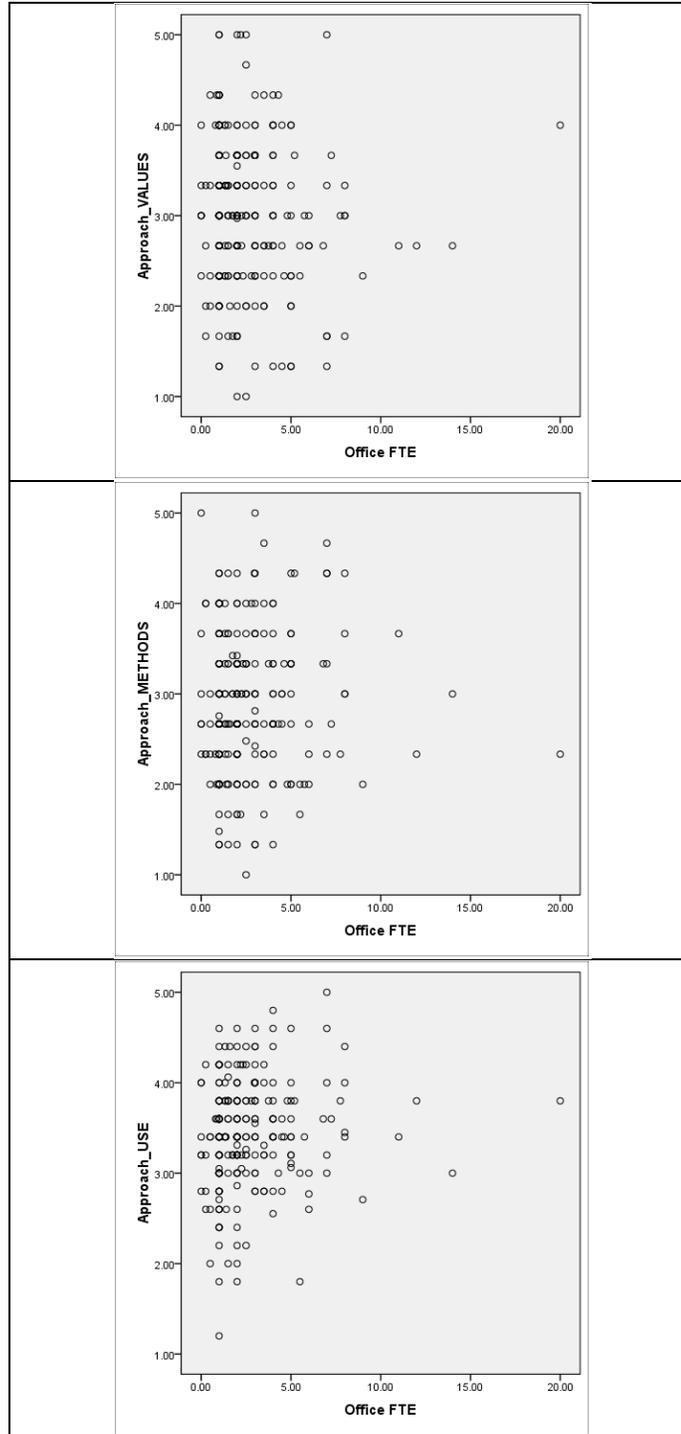




### Assessment/Evaluation Practices Tree



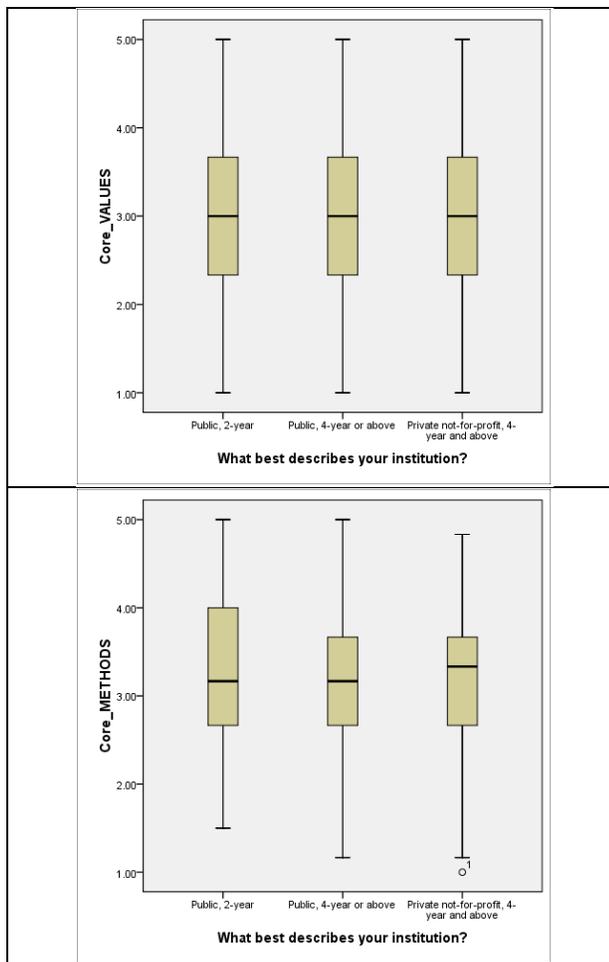
### Approaches to Conducting IR Tree

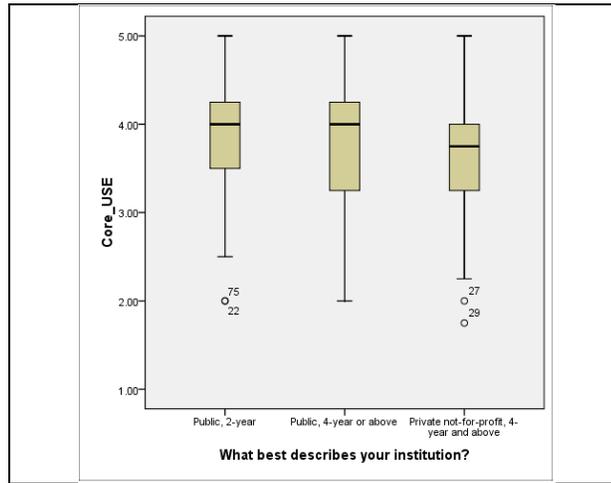


## APPENDIX D

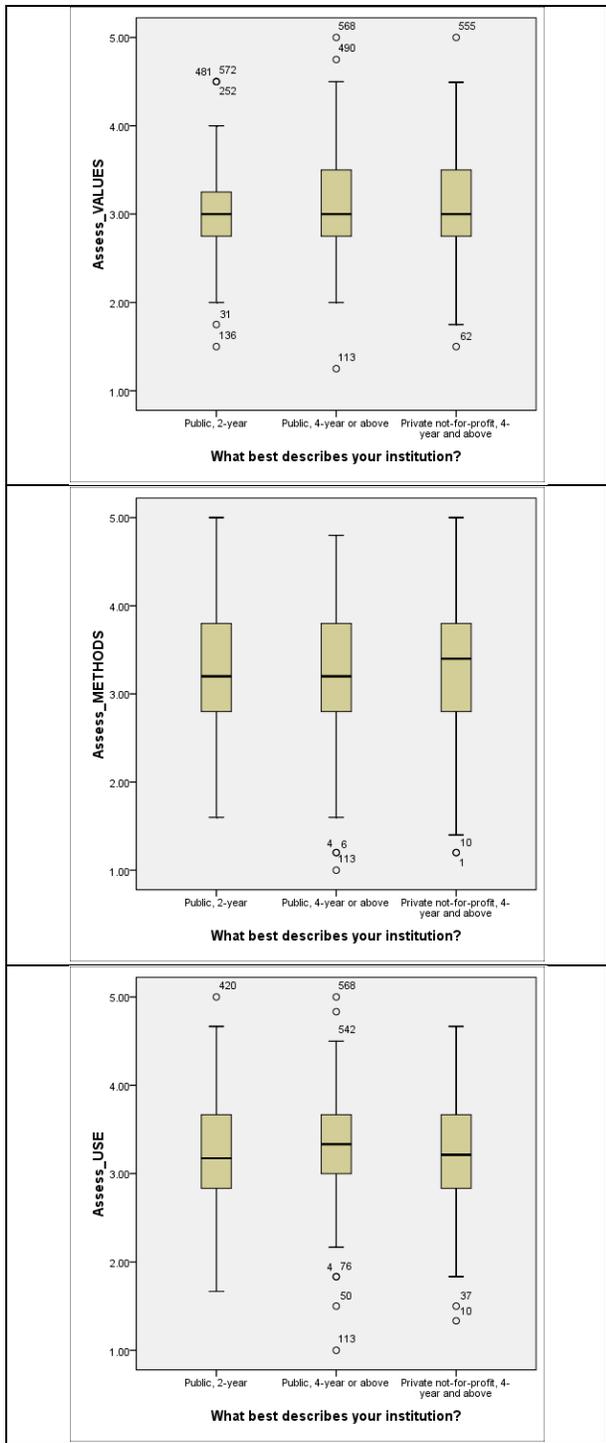
## BOX PLOTS

## CORE IR Reporting Tree





### Assessment/Evaluation Practices Tree



### Approaches to Conducting IR Tree

