



4-2008

Legislative Program Evaluation Conducted by State Legislatures in the United States

John S. Risley
Western Michigan University

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the Policy Design, Analysis, and Evaluation Commons, Public Administration Commons, and the Public Affairs Commons

Recommended Citation

Risley, John S., "Legislative Program Evaluation Conducted by State Legislatures in the United States" (2008). *Dissertations*. 809.

<https://scholarworks.wmich.edu/dissertations/809>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



LEGISLATIVE PROGRAM EVALUATION CONDUCTED BY STATE
LEGISLATURES IN THE UNITED STATES

by

John S. Risley

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Interdisciplinary Ph.D. in Evaluation
Dr. Gary Wegenke, Advisor

Western Michigan University
Kalamazoo, Michigan
April 2008

LEGISLATIVE PROGRAM EVALUATION CONDUCTED BY STATE
LEGISLATURES IN THE UNITED STATES

John S. Risley, Ph.D.

Western Michigan University, 2008

This study examines how U.S. state legislative staffs conduct evaluations. The study addresses the ubiquity of state legislative program evaluation (LPE) units, the standards those units follow, the recommendations that LPE reports proffer, and the quality of the reports on several criteria. The study also addresses the feasibility of using metaevaluation to evaluate a large number of reports using solely the information contained in the reports.

The study uses metaevaluation criteria developed by combining aspects of, primarily, the Generally Accepted Government Auditing Standards (GAGAS) for performance audits, the Joint Committee's Program Evaluation Standards (PES), and, secondarily, Scriven's Key Evaluation Checklist. In the process of developing the metaevaluation criteria the GAGAS and the PES are closely compared. The criteria were applied to a random sample of 100 of the 1,911 LPE reports published by state LPE units from 2001 through 2005.

The study finds that state LPE units, and consequently the reports they produce, are overwhelmingly more connected to performance auditing and the GAGAS than to evaluation and evaluation standards. The metaevaluation criterion on which the LPE reports varied most was the comparisons criterion. Roughly a third of

all LPE reports were graded excellent or good, another third fair, and the final third poor—reflecting no mention of comparisons in the report. Evaluations were more likely to be graded excellent or good on this criterion than were performance audits.

This study also seeks to test a methodological model—that of using metaevaluation to examine a large number of reports. The results from this attempt are mixed. Using metaevaluation in this way can determine the specific areas where evaluation reports are excelling or failing. However, accurately and fairly evaluating reports solely from the report itself presents some major problems. Among these problems are the inability to check both the accuracy of most data collected and the propriety of techniques used to collect data from human subjects. Nevertheless, we can formulate important conclusions including how well LPE reports use comparative studies when reaching their conclusions, how focused the reports are on goals and objectives, and how closely the reports follow established professional standards.

Copyright by
John S. Risley
2008

UMI Number: 3303474

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3303474

Copyright 2008 by ProQuest LLC.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGMENTS

I would like to thank the members of my committee. Michael Scriven always offered sharp critiques and was generous with his time, even after his retirement from the program. James Sanders and Chris Coryn both were very kind to donate their time and offered very insightful comments that greatly improved early drafts.

I would also like to thank several other faculty members at Western. Jane Davidson gained university approval of this program and so ably led it for my first two years. Matthew Mingus served as my advisory committee chair. He gave much time and helpful guidance to a student from a program outside his home department. Gary Miron served on my advisory committee and brought me to the Evaluation Center and to the field of evaluation.

Thanks to my fellow students in the interdisciplinary evaluation program. It has been exciting to be a part of such a talented and diverse group. I am especially grateful to my colleagues from the first cohort—Thomaz Chianca, Cristian Gugui, Paul Lamphear, and Daniela Schröter.

My work examined the work of others—staff members at state legislative program evaluation units. I thank them for their professionalism and public service.

Finally, thanks to my family. My mother's example inspired me when it seemed pursuing my education, working, and raising a family were too much to handle at the same time. My wife, Julie, supported me throughout and always believed in me more than I believed in myself. Our children, Zachary and Kate, constantly reminded me what was most important.

John S. Risley

TABLE OF CONTENTS

| | |
|---|----|
| ACKNOWLEDGMENTS | ii |
| LIST OF TABLES | vi |
| CHAPTER | |
| I. INTRODUCTION | 1 |
| II. LITERATURE REVIEW | 4 |
| Terminology | 6 |
| Auditing and Evaluation Compared | 13 |
| National Supreme Audit Institutions | 16 |
| Legislative Program Evaluation and the States | 19 |
| Structure, Growth and Staffing | 20 |
| Methodology: Surveys..... | 29 |
| Methodology: Case Studies..... | 31 |
| Methodology: Directly Examining Reports | 33 |
| Metaevaluation | 35 |
| Background and Comparison of Evaluation and Performance Audit Standards | 50 |
| Generally Accepted Government Auditing Standards for Performance Auditing..... | 51 |
| Program Evaluation Standards | 54 |
| Conclusions: PES and GAGAS Standards..... | 81 |
| Summary and Conclusions | 82 |
| III. METHODOLOGY | 84 |

Table of Contents—continued

| | |
|--|-----|
| CHAPTER | |
| Legislative Program Evaluation Units | 85 |
| Non-legislative Program Evaluation Units | 88 |
| Report Recommendations | 89 |
| Louisiana Subsample | 91 |
| Metaevaluation Criteria..... | 91 |
| Preliminaries | 97 |
| Methodology..... | 102 |
| Results | 107 |
| IV. FINDINGS..... | 115 |
| LPE Metaevaluations | 115 |
| Preliminaries..... | 116 |
| Methodology..... | 117 |
| Results | 119 |
| Discussion..... | 119 |
| Comparisons | 122 |
| Legislative Versus Non-legislative Program Evaluation Units..... | 125 |
| Preliminaries..... | 125 |
| Methodology..... | 126 |
| Results | 127 |
| Report Recommendations | 129 |
| V. CONCLUSIONS | 133 |

Table of Contents—continued

CHAPTER

| | |
|---|-----|
| LPE Units and Reports..... | 134 |
| Metaevaluation..... | 137 |
| Key Criteria for Metaevaluating LPE Reports | 140 |
| Directions for Further Research | 141 |
| Comparisons with Previous Literature..... | 143 |

| | |
|------------------|-----|
| REFERENCES | 145 |
|------------------|-----|

APPENDICES

| | |
|--|-----|
| A. Metaevaluation Criteria | 152 |
| B. List of LPE Sample Reports | 159 |
| C. List of Non-LPE Sample Reports | 167 |
| D. List of Louisiana Sample Reports..... | 169 |

LIST OF TABLES

| | |
|--|-----|
| 1. Metaevaluation Literature: Multiple Studies | 37 |
| 2. Program Evaluation Standards Versus GAGAS..... | 55 |
| 3. Reports Sample and Population..... | 87 |
| 4. State Non-legislative Program Evaluation Units..... | 88 |
| 5. State Non-legislative Program Evaluation Reports | 89 |
| 6. Metaevaluation Criteria and Sources..... | 93 |
| 7. Confidence in Metaevaluation Criteria..... | 95 |
| 8. LPE Reports Sample: Standards Cited | 116 |
| 9. LPE Metaevaluation Ratings: Preliminaries..... | 117 |
| 10. LPE Metaevaluation Ratings: Methodology..... | 118 |
| 11. LPE Metaevaluation Ratings: Results | 120 |
| 12. LPE Reports Using Comparisons | 123 |
| 13. Comparisons Criterion: Non-LPE Versus LPE Reports..... | 124 |
| 14. Non-LPE Metaevaluation Ratings: Preliminaries | 126 |
| 15. Non-LPE Metaevaluation Ratings: Methodology | 128 |
| 16. Non-LPE Metaevaluation Ratings: Results..... | 129 |
| 17. LPE Reports Sample: Standards Cited | 129 |
| 18. LPE Reports with Recommendations..... | 130 |
| 19. Non-LPE Reports with Recommendations..... | 131 |
| 20. LPE, Non-LPE, and Louisiana Reports: Recommendations | 132 |

CHAPTER I

INTRODUCTION

The main policy-making bodies for state governments in the United States are the respective state legislatures. In the U.S. federal system the states are called on to enact and carry out a myriad of policies and programs from education to health to worker safety. How state legislatures conduct evaluation in both their policy-making and oversight roles would seem a crucial aspect of evaluation work. Evaluation likely influences how states determine policy choices, enact policies from among those choices and decide on the value of those choices once enacted. However, very little has been written about how state legislative staffs conduct evaluations.

There is a body of evaluation literature surrounding national-level evaluation agencies, Supreme Audit Institutions (SAIs), around the world (see Gunvaldsen & Karlsen, 1999; Leeuw, 1992; Ling, 2003; Pollitt & Summa, 1996). At the level of U.S. states the published, academic literature concentrates on format, process, and scope rather than the products—the reports—of the state legislative program evaluation (LPE) units.

This study concentrates on state legislative program evaluation units. While much evaluation work concentrates on the executive or administrative side of evaluation use, comparatively little is written about the evaluations state legislatures commission from their own staffs. Since legislatures are the predominate policy-making and funding bodies, the direction they set on evaluation is extremely important.

The study seeks to contribute to the knowledge of how evaluation is conducted by an important U.S. governmental sector, state legislatures, on a wide range of issues (education, transportation, safety, etc.). It also aims to show how metaevaluation can be used to assess a large number of evaluation reports in a setting, field or topic.

Henry and Mark (2003) discuss the relative dearth of research concerning evaluation itself. They contrast this with the rather extensive literature concerning evaluation findings—what programs work and what do not. They write that not much is known by evaluators about “how many evaluations were completed last year, how much money was spent on evaluation, who did the evaluations, and how they were conducted.” (p. 69) There is a “serious shortage of rigorous, systematic evidence that can guide evaluation or that evaluators can use for self-reflection or for improving their next evaluation.” (p. 69) Henry and Mark note the following:

Prescriptive advice and admonitions about how to do evaluation have been plentiful, filling books, journals, conferences, e-mails, and conversations. But these are generally based on personal experience, observation, and the individual’s sometimes idiosyncratic beliefs and values—and not on carefully gathered evidence that can be described, shared, and critiqued. (p. 70)

This statement applies directly to the legislative program evaluation literature as well as the general evaluation literature.

After reviewing the state of the evaluation literature, Henry and Mark suggest six different areas that are ripe for research on evaluation. Among these six is metaevaluation. They write that metaevaluation research “can be informative about the nature of evaluation practice in specific cases” and that “the greatest benefit of

metaevaluations to the field will come from syntheses across multiple metaevaluations.”
(p. 74)

This study incorporates the metaevaluation item. A base of knowledge about what is currently happening in the legislative program evaluation area is important to inform any future research. The study contributes to the practice of evaluation in two ways. First, the study furthers the knowledge of how evaluation is conducted by the main U.S. governmental sector on a wide range of issues (education, transportation, safety, etc.). Second, this study contributes to the literature concerning metaevaluation—specifically the feasibility of conducting metaevaluations using solely the evaluation report. It contributes to the theory of evaluation by showing how evaluation works in combination with (and sometimes within) other fields: legislative governance and auditing.

CHAPTER II

LITERATURE REVIEW

This review of current literature surrounding state legislative program evaluation (LPE) will be divided into five sections. First, a section on terminology will examine the various terms and descriptions used to explain what state legislatures do to evaluate programs, policies and organizations. Many different terms have been used with some general areas of agreement appearing over time. However, there are still many activities that appear to be the same yet are designated by different terms. Second, a look at the literature regarding the separate fields of auditing and evaluation, how they have come together in many instances, and how certain types of auditing (performance auditing, program auditing, economy and efficiency auditing) are strikingly similar to evaluation. Third, a brief look at the literature addressing auditing at the national government level. This literature is addressed to help inform the examination of LPE at the state level. Fourth, an analysis of the literature specifically addressing LPE in state legislatures. Themes in this section include the growth of LPE in the states, different types of configurations for LPE offices/organizations, how LPE can best ensure use of their findings, and future prospects for LPE in the states. The fifth and final section will examine the literature from the methodological perspective. Authors collected original data through three main methods: surveys of LPE units and other interested stakeholders, case studies of LPE in various organizations, and a direct examination of the product of

LPE in the states—the evaluation reports themselves. It is this third method—the one that will be employed in this study—that has barely been utilized at the state level.

Broadly, the literature in three of the five areas is much more descriptive than analytical. The literature in the first section mainly describes the current typology surrounding performance auditing and legislative program evaluation. Similarly, the literature concerning the separate fields of auditing and evaluation describes differences from a theoretical level. The literature specifically addressing LPE in state legislatures is also generally descriptive—addressing topics such as the educational background of state LPE staff members, the activities undertaken by LPE units, and other descriptive issues.

The two areas that show a slightly more analytic bent address auditing at the national government level and the methodological strategies employed in some of the literature. The national level auditing literature offers a glimpse of what the state LPE literature could become, given more attention by scholars. The fifth area examined—the methodological perspective—shows surveys and case studies generally employed for descriptive rather than analytical purposes.

After reviewing the LPE literature, attention focuses on the literature concerning metaevaluation. This includes a discussion of the published literature that uses metaevaluation across multiple evaluations (focusing on the content area, the standards used, and the number of evaluations included) and the body of literature addressing the theoretical aspects of metaevaluation. This literature addresses questions such as how to conduct a metaevaluation and when metaevaluation is needed.

The metaevaluation literature is helpful because it shows the different ways metaevaluation is used to examine evaluations. This literature informs subsequent

metaevaluative studies much more than descriptions of state LPE units inform subsequent studies of that subject.

Terminology

The activities undertaken by state LPE units share many similarities yet are designated by many different terms. This study adopts the term “legislative program evaluation” to refer to the variety of evaluative activities carried out by these units. LPE is the term used by the National Conference of State Legislatures’ staff section, the largest membership group of these organizations. Though many states use terms such as performance audit and program audit to describe these practices and offices, the evaluative aspect is present in the work performed. The discussion below follows the literature chronologically to explore the emergence and disappearance of different terms.

Knighton (1967), being one of the earliest attempts to look at performance auditing in state legislatures, is a logical place to start in an examination of the terminology used to describe this issue. Knighton adopted the term *performance post audit*, which he defined as

. . . an independent examination, conducted by the Auditor General, for the purpose of providing the legislature with an evaluation and report of the manner in which administrators of the agencies and departments of the state have discharged their responsibilities to faithfully, efficiently, and effectively administer the programs of the state. Faithfulness refers to whether or not programs have been administered in accordance with promises made to the legislature and the expressions of legislative will. Effectiveness refers to whether or not planned program objectives have been achieved. Efficiency refers to whether or not program accomplishment has been achieved by using the least cost combination of resources and with a minimum of waste. (p. 1)

He sets out three criteria that are echoed in descriptions of legislative program evaluation to come: faithfulness, efficiency and effectiveness. He uses the term evaluation to describe what auditors do for the legislature. While Knighton does broach the topic of evaluating legislative goals, through his faithfulness criterion, conspicuously absent is any mention of examining non-goal related outcomes, or side-effects. This is a key aspect of evaluation (Scriven, 2006) but one that is often, though not always, overlooked in discussions of LPE.

One of the earliest uses of the term *legislative program evaluation* is by Chadwin (1974). He cites different names that various agencies, offices, and units in the states use to describe LPE activities—*program auditing*, *program review*, *performance post-auditing*, *legislative oversight*, and *effectiveness auditing*. He concludes that while there is different terminology, and some degree of different approaches employed, the objective of each LPE unit is to “determine the effectiveness—the results—of public programs” (p. 1). The heart of Chadwin’s definition is the concept of program evaluation which, in the state legislative context, differs from traditional staff oversight work in that it 1) is concerned with programs and therefore cuts across organizational structures, government departments, and agencies, 2) it “emphasizes output and impact as well as input and process” (p. 1), and 3) it entails much more intensive data collection and analysis than traditional legislative staff work. Chadwin’s use of the term that has come to be accepted in the field—legislative program evaluation—and his differentiation between LPE and traditional legislative oversight was an important early advance in identifying and establishing LPE.

Writing the same year as Chadwin, Brown and Pethel (1974) address many of the same, or similar, terms but in greater depth. They adopt the term *program evaluation*, stating that “as an institutional commitment of state legislatures [it] is a recent phenomenon” (p. 318). They then use the term *legislative evaluation*, writing that it is “a departure from the traditional concern of public administration with the executive and, as such, needs explanation, description, and differentiation from other forms of evaluation” (p. 318). They contend that the “legislative role in evaluation is not sufficiently well developed for us to speak in terms of generally accepted procedures, methods or style” (p. 318). They employ a variety of terms—*legislative post-audit*, *performance audit*, and *legislative performance audit*—interchangeably and contend these terms are becoming generally accepted with a majority of state legislatures devoting a legislative agency or unit to these tasks. They show the close linkage between auditing and evaluation by describing legislative performance audits as producing evaluation reports. They also interchangeably use the terms *legislative evaluation organization* and *state legislative audit groups* to describe the various state legislative agencies and units that perform these activities. Throughout this study the term *LPE unit* is used.

Brown and Pethel contend that working definitions in the post-audit field are emerging for performance audits and management audits. A performance audit is “an evaluation of the effectiveness of governmental operations, programs, and organizations to determine accomplishment of goals and objectives” (p. 319). A management audit is defined as “an evaluation of the efficiency of governmental operations, programs, and organizations with special attention to administrative policies and practices” (p. 319). While these definitions are emerging, state LPE units do not agree on specific definitions.

While touching on many of Chadwin's points, Brown and Pethel continually conflate different terms. This lessens their influence in defining distinctions between LPE and other activities.

In another mixture of auditing and evaluation Brown (1976) uses the term *program evaluation audits*. He believed there was, and would continue to be, a blurring between the activities of auditors and those of evaluators. He cited state-level auditors being asked frequently by legislators if they could take on more program evaluation work. Since many states, at this time, had financial audit units available it seemed easy to add more responsibilities to these existing units, rather than create a new, separate, evaluation unit. However, he cautioned that such an approach would produce "a hybrid which is unsatisfactory from all points of view" as the new program evaluation work would leave less time for the important financial audit work to which legislators had become accustomed. Brown's main contribution is identifying the slow expansion of auditors' duties to include program evaluation. A more quantitative examination of this trend, through either a survey or an examination of reports issued by state-level auditors, would have added value to his contribution.

Writing a few years later Brown (1979) changes slightly from his previous term and adopts *audit-evaluation* to describe "at a minimum, program or performance auditing and program evaluation" (p. VIII). He points out that states with offices designated *post audit* usually use the term *program auditing* while those state offices with different designations use the term *program evaluation*. He points out that another, slightly less commonly used, term is *program review*. Though this term is not as common, it is the one he uses in the title of this book—*The Effectiveness of Legislative Program Review*.

As with his earlier work, this book would have benefited from a more rigorous examination of the various state audit offices.

Craft (1979) also uses the term audit-evaluation when referring state legislative oversight work. He defines *legislative oversight* very narrowly to mean “published reports written by State legislative staffs whose primary focus is to evaluate the efficiency and/or effectiveness of government programs” (p. 22). Again, the terms audit and evaluation are combined.

Jones (1987) discusses state legislative activities that are variously called *program evaluation*, *performance auditing*, *program review*, and *sunset*. These all have a “common mission: to review, analyze and asses how state agencies and programs have been working” (p. 20). He refers to the state legislative organizations as evaluation units and program evaluation units. The term sunset here refers to a reform instituted in the states in the 1970s where the legislative authorization of programs and agencies was granted with a sunset, or expiration, date after which the program or agency would either need to be re-authorized by the legislature or terminated (see Adams & Sherman, 1978).

Greathouse and Funkhouser (1988) cite Knighton (1967) as the basis for their definition of performance auditing: “an independent examination for the purpose of reporting on the extent to which a state agency is faithfully, efficiently, and effectively carrying out the programs for which it is responsible” (p. 57). The authors believe that performance auditing is not program evaluation. In their view program evaluation

... seems to be used primarily by academics in education and psychology to describe an activity more related to scholarly research than to legislative oversight. Program evaluation seems to denote an activity that is largely sympathetic, nonjudgmental, and collaborative in relation to the entity being evaluated. (p. 58)

Conversely, performance auditing “produces sharper, more focused and more critical conclusions on major issues than either program evaluation or other forms of nonfinancial auditing” (p. 58). This is not a view of evaluation shared by many in the evaluation field.

Davis (1990) makes comparisons between the two traditions of performance auditing and evaluation research. Davis adopts the Generally Accepted Government Auditing Standards (General Accounting Office) definition of *performance auditing*, a technique that involves “a determination of the economy, efficiency, and effectiveness of government organizations, programs, activities and functions, in addition to their compliance with laws and regulations” (p. 35). He adopts the Rossi and Freeman (1985) definition of evaluation research: “the systematic application of social [research] procedures in assessing the conceptualization and design, implementation and utility of social intervention programs” (p. 35). Davis is one of the few authors to adopt definitions from authoritative sources, be they regulatory agencies or scholars from another field.

Sedgwick (1993), writing of the usage in Australian government, defines *performance auditing* as “concerned with auditing qualitative and quantitative measures in the public sector with the aim of improving program operational efficiency as well as administrative effectiveness” (p. 148). It is also known as *efficiency auditing*. Sedgwick contends evaluation addresses these same areas but can, and should, also focus on the effectiveness and appropriateness of programs and the policies that frame them. This additional focus of evaluation is what essentially differentiates it from performance auditing. Sedgwick is one of the few authors to broach this issue of evaluation’s mandate to address program appropriateness.

Brooks (1996) argues that a blended approach encompassing auditing and evaluation has developed, largely in response to the needs of state legislators. Like Davis (1990) he uses the GAO's definition of *performance auditing*. Though he notes that *program evaluation* is much more difficult to define because it lacks the formal organizational underpinnings, he also adopts the Rossi and Freeman (1995) definition of evaluation.

Barzeley (1997) sees a struggle between those who would define performance auditing more along the lines of traditional financial auditing and those who define it more like evaluation (Chelimsky, 1996; Rist, 1989). Barzeley takes the view that "performance audits are not a type of audit; they are evaluations" (p. 237). He points out differences between performance auditing, evaluation and auditing and tries to show performance auditing as a concept rather than just give a definition. He addresses how European countries define performance audit and the activities that they consider encompassing performance audits. This helpful venture beyond the borders of North America is conspicuously absent in most of the other literature concerning performance auditing.

Throughout the literature on terminology auditing and evaluation are often combined and sometimes conflated. In general, both terms and their extensions (for auditing these include performance auditing, program auditing and efficiency auditing, for evaluation they include program evaluation, legislative program evaluation and evaluation research) are defined as assessing the conduct, results and impacts of government programs, policies and agencies. The ubiquity of the words *evaluation* and

evaluate in definitions provided for performance and program auditing further illustrates how closely related the fields of performance auditing and evaluation are.

Auditing and Evaluation Compared

There is a limited amount of literature addressing the similarities and differences of auditing and evaluation. The papers in this area come almost equally from the accounting/auditing side and the evaluation/public administration side. Beginning in the 1960s and 1970s public administrators and auditors began to link a number of government performance improvement movements with the fields of auditing and program evaluation. While some address a comparison of auditing as a whole to evaluation, others concentrate on performance auditing and its relationship to evaluation.

Chelimsky (1985), writing in the journal *Evaluation Review*, provides the most comprehensive and informative discussion of the origins, differences and similarities of auditing and evaluation. Auditing developed, historically, “as a procedure for detailed checking” (p. 485) concerned most with verification. Evaluation, conversely, is more concerned with the “actual effects of government programs and policies” (p. 486). While both auditing and program evaluation address normative questions descriptive questions are routine in program evaluation but rare in auditing.

Davis (1990) makes comparisons between performance auditing, not auditing more broadly, and evaluation research on the following four dimensions: variables emphasized, independence, quality control, and professionalism. Performance auditors emphasize “management control” more and program impacts less while evaluators do the

opposite. Auditors are more concerned with legal and procedural compliance issues than are evaluators.

Pollitt and Summa (1996) contend that there is an absence of empirical investigation matching audits and evaluations to discover their similarities and differences (though the authors point to such a project they are working on). They largely draw on the literature about Supreme Audit Institutions (SAIs) in Europe for their perspectives on performance audits. They cite Chelimsky (1985) as the definitive paper on the subject but point out that her focus was on traditional financial auditing, not performance auditing. Most of their comments about evaluation units do not envision an evaluation unit within a legislative body. For example, they contend that SAIs have more authority to get their reports heard among legislators than do evaluators. This clearly overlooks evaluators who work for legislative bodies.

Rist (1989) largely echoes Chelimsky (1985) on the differences between auditing and evaluation. His main contribution is his discovery that 70 percent of employees at the U.S. Congress' General Accounting Office (GAO, now called the Government Accountability Office) are designated as evaluators while only a minority is classified as accountants.

Newcomer (1994) surveyed United States federal government Inspectors General and evaluation offices regarding their audit and evaluation activities. She reports that both of these organizations focus on program process and implementation rather than on impacts. Independence is more important to auditors than to evaluators. Evaluators are much more likely to invite program staff to be on evaluation teams and more generally involve them in the evaluation process. This use of a survey to discover the views of

auditors and evaluators at the federal level in the United States is unique. Several survey have been conducted at the state level (Botner, 1986; Mohan, 1997; NPLES, 2004; Schwartz & Mayne, 2005; Stutzman & Rainey, 1981; VanLandingham, Kinton, Boerner, & Zelio, 2000; Walton & Brown, 1990).

Sedgwick (1993) contends that performance auditing and evaluation address similar questions but evaluation can, and should, also focus on the effectiveness and appropriateness of programs and the policies that frame them. This additional focus of evaluation is what essentially differentiates it from performance auditing. Other differences between performance auditing and evaluation include: performance auditing is more concerned with compliance and operational efficiency, performance auditors (in Australia) do not question policy outcomes while evaluators do, evaluation relies on a much broader base of data, and performance auditing relies on authoritative/official data while evaluation collects additional data. The similarities between performance auditing and evaluation include: both are fundamental parts of the accountability chain, both contribute to better program management, both rely on objectivity, integrity and professionalism, and both cover a diverse range of activities.

Brown (1976) contends that the presence of “single purpose legislative program evaluation staffs in a number of states” (p. 85) has caused legislators to ask state auditors about doing more performance work, program evaluation work and program audit work. As discussed above in the terminology section, he uses the term *program evaluation audits* because he believes there is, and will continue to be, a blurring between the activities of auditors and those of evaluators. However, he believes a traditional audit agency, unchanged, cannot perform program evaluation audits. His experience in the

Kansas Legislative Division of Post Audit is that they have two divisions, one for financial-compliance audits and one for program evaluation audits. Brown quotes Mark Chadwin (from personal correspondence) as saying “program evaluators generally will go further than auditors (1) to construct program objectives where statutory and documentary sources are vague and (2) to question existing program objectives if their research raises doubts about them” (p. 86).

The limited literature comparing auditing and evaluation does highlight the main agreement on differences between the two fields. Auditing, even when confined to performance auditing, is more concerned with process and adherence to legislative directions and goals. In contrast, evaluation is more concerned with outcomes and impacts. While evaluation is also sensitive to legislative goals it can often question the goals themselves.

National Supreme Audit Institutions

At the national level, government performance auditing is carried out by Supreme Audit Institutions (SAIs). The American SAI is the Government Accountability Office (formerly the General Accounting Office). There is a voluminous literature about SAIs and their traditional or financial auditing duties. The literature concerning their performance auditing duties is less extensive (internationally, performance auditing is often referred to as economy and efficiency auditing or value-for-money auditing). A review of the SAI literature on performance auditing gives some links to how performance auditing works at the state legislative level in the U.S.

Roberts and Pollitt (1994) present a case study of a value-for-money (VFM) audit completed by the British National Audit Office. The authors view the VFM as a step beyond traditional auditing but still not a “full-blooded evaluation” (p. 527). One way in which it falls short of a program evaluation is that VFMs in Britain are not permitted to “question the appropriateness of government policy objectives themselves” (p. 545). Roberts and Pollitt see three limitations of the VFMs: (1) reports rarely specify methods for improvement; (2) scope of work is primarily financial; and (3) dependence on parliament means they are constrained by time.

Barzeley (1997) offers a basic primer on how European countries define performance auditing and the activities that they consider encompassing performance audits. He argues that audit bodies charged with conducting performance audits face a decision of whether to adopt a “machine-like or rule-governed” approach or a “more particularistic style of ideal-typical professional bureaucracies” (p. 255). Most choose the rule-governed approach in an attempt to maintain their legitimacy.

Two papers about SAOs (Walker, 1985, and Schwartz, 1999) examine the actual reports of SAIs to determine what types of issues they are looking at and how effective they are.

One of the most methodologically sound papers that concentrates on the American SAI, the GAO, is Walker’s (1985). Walker states that the GAO’s program evaluation efforts serve Congress by providing oversight of public policy, a bludgeon to coax agencies into pursuing activities preferred by Congressmen, and a tool for electoral activities. Methodologically, Walker reviewed a broad sample of reports from the GAO (35 in their entirety and 100 summaries) and determined that the GAO reports encourage

growth in the size and scope of the state. He reviewed financial, management and program audits/evaluations. “Most members of Congress do not find that pursuing the auditor’s norms of economy and efficiency contribute to their quest for reelection, to gaining influence in their chamber or their committee, or to achieving desired policy outcomes” (p. 365). Walker cites Fenno’s (1974) classic work on congressmen in committee as identifying these typical desires. Addressing the impacts of performance audit and evaluation work by the GAO, Walker contends that GAO reports are a minor claimant on the attention of public managers. Walker believes auditors’ values (essentially economy and efficiency) are inconsistent with democratic values in the United States (redundancy, federalism, separated powers, and various other checks on legislative and executive power). While the GAO can act as a watchdog and “stop people from doing some bad things” (p. 365) it doesn’t have the authority/power to force change.

Schwartz (1999) examined “effectiveness audits or evaluations” (p. 513) conducted by SAIs in six different countries to see how they really address effectiveness in light of the political difficulties of doing so. He found that “most state auditors have not met the challenge non-politicized, professional and objective outcome evaluations in place of agency-based evaluations which are perceived to be infected by political interests” (p. 522). He helpfully broaches the political aspects of a state auditor’s job in his conclusion that the main reason for this is the auditors’ desire to avoid the political dangers inherent in questioning the effectiveness of programs.

Schwartz and Mayne (2005) look at how government organizations in different countries evaluate their evaluation, performance auditing and performance reporting information. They cite Schwartz (1999) as a previous report examining or metaevaluating

audit reports. Their methodology is basically a survey of SAIs asking them what type of quality assurance systems they have in place. Unfortunately, the authors did not look directly at the quality of the work. The surveys show that political concerns are one factor inhibiting quality assurance activities in this area.

Song (2003) addresses the supreme audit institutions in South Korea, the Board of Audit and Inspection (BAI). He shows how they are moving from performance audits to program evaluations. As the BAI sees it, performance audits concern “the extent to which the expenditure and implied actions have minimized the inputs for stipulated outputs (economy), or maximized the outputs from the inputs used (efficiency), or maximized the desired products or impacts of programs (effectiveness)” (p. 9). Song contends that one of the main things the BAI needs to do when moving towards more program evaluation is to expand their training unit to include research and evaluation methods as well as accounting and audit methods. Also, the BAI needs to recruit more staff with skills in “data collection, statistical research and policy analysis” (p. 13). These may be sound recommendations, but Song presents no evidence that current BAI staff lack these skills.

Legislative Program Evaluation and the States

The literature on legislative program evaluation began with Knighton in 1967. The early seventies through the mid-1980s saw a fair amount of interest in the issue, as the number of states performing legislative program evaluation grew. Since 2000 there has been very little addressing LPE. Almost no literature looks directly at the products—the reports themselves—of legislative program evaluation in American states. Instead

much is written about the growth of LPE units and LPE activity in the states, how to structure a legislative program evaluation unit, and what type of people work in this field. Methodologically, many papers use surveys to address unit structure, growth and activities, others use case studies about different audit/evaluation reports, and finally two papers look directly at a sample of evaluation-audit reports.

Structure, Growth and Staffing

Knighton (1967) basically defined performance post audit. This was the first real academic look into this at the state level. He focused on the state of Michigan and its creation of an auditor general office. Knighton stresses the importance of seating the auditor under the control of the legislature and discusses the various pros and cons when this reform was adopted in the 1962 Michigan constitution.

Crane (1972) lists information such as budgets, staffing levels, and staff member salaries for legislative services agencies in each state. Crane's work is cited in some papers (Berry, Turcotte, & Latham, 2002; Studer, Spitz, & Burt, 1981) to show the growth of legislative audit/evaluation offices in the states.

Brown and Pethel (1974) provide a breakdown of some early legislative evaluation/performance audit units in state legislatures. They adopt the term legislative performance auditing but fail to clearly define it and use it interchangeably with legislative evaluation and evaluation research. Brown and Pethel conclude that "the legislative role in evaluation is not sufficiently well developed for us to speak in terms of generally accepted procedures, methods, or style. But, there is an inkling of evidence that

its precepts differ from standard public administration principles of evaluation research” (p. 318).

Chadwin (1974) writes that in 1970 no state legislature had a full time staff responsible for evaluation of program effectiveness. By 1974 over a dozen did (committees, commissions or auditors offices). He cites different names for these activities, such as: *program auditing*, *program review*, *performance post-auditing*, *legislative oversight*, and *effectiveness auditing*. Chadwin list the following reasons for the increase in legislative program evaluation offices: an increased workload for legislatures, federal decentralization efforts, heightened executive-legislative tension, citizen pressure, and the availability of trained personnel. Legislative program evaluation units are usually organized in one of three ways: under the control of an independent commission (or committee), as core legislative staff or an auxiliary unit to the core legislative staff, under the control of a legislative auditor’s office.

Brown (1976) contends that the presence of “single purpose legislative program evaluation staffs in a number of states” (p. 85) has caused legislators to ask state auditors about doing more performance work, program evaluation work and program audit work. As discussed above in the terminology section, Brown uses the term *program evaluation audits* because he believes there is, and will continue to be, a blurring between the activities of auditors and those of evaluators. However, he believes a traditional audit agency, unchanged, can perform program evaluation audits. His experience in the Kansas Legislative Division of Post Audit is that they have two divisions, one for financial-compliance audits and one for program evaluation audits. Brown quotes Mark Chadwin (from personal correspondence) as saying “program evaluators generally will go further

than auditors (1) to construct program objectives where statutory and documentary sources are vague and (2) to question existing program objectives if their research raises doubts about them” (p. 86).

Craft (1979) points out that legislative auditor offices often have a program audit division and a (much larger) financial audit division. Also some legislative fiscal staff members and general legislative research staff members who serve standing committees conduct performance audits. He offers a comprehensive view of the state of LPE in the late 1970s. By the middle of 1978 the Legislative Program Evaluation Section Clearinghouse at the Eagleton Institute of Politics at Rutgers had received 477 “state legislative program and management audit-evaluations” (p. 23). Of these, 39 percent were produced by legislative auditors, 35 percent produced by special purpose evaluation staffs, 14 percent by fiscal offices, 9 percent by general research staffs, and 3 percent by individual committee staffs (p. 23). Almost all of the oversight staffs report to joint legislative committees or commissions while all the special purpose evaluation staffs report to joint legislative committees. In 13 of the 17 states studied staff members must receive formal approval of a topic from their oversight committee/commission before starting an audit-evaluation. Recommendations in reports are sometimes revised by the legislators on oversight committees/commissions before a report is released. A few states do not provide recommendations in their reports. Several states produce annual reports to the legislature that include follow up on recommendations from reports. While the number of states with these types of organizations/staffs is increasing there are also many instances of program audit/evaluation sections closing or ceasing program evaluation activities. Audit-evaluation staffs do not like taking on the sunset review duties. They feel

these are more routine/unimportant regulatory bodies and these duties take time from their other responsibilities. “Extensive program and management audit-evaluation work by the State legislative staffs is less than 5 years old in most States. The next decade will see a considerable amount of turmoil before formal legislative oversight becomes a normal, accepted practice” (p. 26).

Brown (1979) uses the term “audit-evaluation” to describe “at a minimum, program or performance auditing and program evaluation” (p. VIII). He points out that states with offices designated *post audit* usually use the term *program auditing* while those in offices with different designations use the term *program evaluation*. Others use *program review*. The development of these offices has followed five stages: 1) recognition, 2) allocation of resources, 3) performance, 4) utilization, and 5) impact. Most of Brown’s book is devoted to case studies of legislative audit-evaluation reports from various states. These case studies allow Brown to uncover interesting conclusions about report implementation. Seven factors affect report implementation in these states: 1) whether a program or an agency was reviewed, 2) whether the legislature approved the topics, 3) the legislature’s level of interest or concern, 4) executive officials’ attitudes toward the reports, 5) the nature to the report findings and recommendations, 6) the timing of the reports, 7) the media coverage generated by the reports.

Studer, Spitz, and Burt (1981) all served in the Minnesota Legislative Auditors Office. They argue that legislative program evaluation can only be effective if: 1) legislators and executive branch officials have reasonable expectations and 2) systematic efforts are made to ensure results are used. They cite the New York Legislative Commission on Expenditure Review as the first state legislative organization for

oversight, created in 1969. Within five years of that more than 12 states had begun related activities. At the time of their writing (1981) approximately 40 state legislatures have the “capacity to conduct formal oversight or evaluation studies” (p. 168). LPE staff must be concerned with getting the legislature to use their work in the budget-policy process and getting the executive branch to use their work to implement program changes. Three things that appear to enhance evaluation’s usefulness are: 1) placement of the evaluation unit, 2) choice of evaluation topics, 3) working within the legislative process. The authors identify these as places where legislative program evaluation units exist: within a traditional financial audit organization, within a fiscal bureau or appropriations committee, within a legislative research unit, and within a standing legislative committee. When choosing an evaluation topic, the authors argue that the following criteria should be used: well-defined use, good policy relevance, a clear program with an existing track record (not new), a stable program (not a lot of senior staff turnover), cost effectiveness (would the cost of an evaluation be greater than the benefits derived from the evaluation), appropriate timing (completed in time to inform budget cycle, legislative session). The most insightful of these criteria is the issue of cost-effectiveness. Unfortunately, no other authors address this issue of weighing an evaluation’s potential benefits against the cost of conducting the evaluation.

Green (1984) writes at a time when legislative performance auditing in Oregon was limited to sunset reviews of some agencies, but when an interim study on expanded program evaluation for the legislature was being conducted. He foresees problems if legislative evaluations are conducted by fiscal committee (such as the ways and means

committee or budget committee) staff because they do not have a broad enough view of policy because they are mostly concerned with spending and taxation issues.

Jones (1987) reports that legislatures in more than 40 states have established some type of mechanism to assess the performance of state agencies and programs. These activities are variously called program evaluation, performance auditing, program review, and sunset. Evaluation units were first established in legislatures in Hawai'i and New York in the late 1960s and adopted by most other states in the 1970s. Nineteen legislatures have created a program evaluation unit within an existing audit agency. 11 legislatures have created a separate program evaluation unit (e.g. NY Legislative Commission on Expenditure Review), and 16 states have assigned program evaluation duties to another legislative agency (in 10 part of a legislative fiscal agency, and in the others part of a broader legislative service agency). Jones quotes Alan Rosenthal (p. 23) as predicting legislatures will use these evaluation units for longer term analysis of policy alternatives.

Greathouse and Funkhouser (1988) state that 33 states have formal audit or evaluator functions, 13 of these are located in a legislative office. The states that call their work program evaluation do not follow the generally accepted government auditing standards (General Accounting Office). The authors see the GAGAS as workable for financial auditing but not for nonfinancial auditing. They argue that the nonfinancial audit and evaluation work being done in state audit organizations in accordance with the GAGAS is called performance auditing. Greathouse and Funkhouser report that no state audit or evaluation unit uses the terms *economy and efficiency auditing* or *program results auditing* (though the GAO does use these terms). They criticize the GAO

standards for conflating economy and efficiency and for not linking performance outputs with the costs of obtaining them. The authors believe that performance auditing is not program evaluation. In their view program evaluation “seems to be used primarily by academics in education and psychology to describe an activity more related to scholarly research than to legislative oversight. Program evaluation seems to denote an activity that is largely sympathetic, nonjudgmental, and collaborative in relation to the entity being evaluated” (p. 58). “Performance auditing produces sharper, more focused and more critical conclusions on major issues than either program evaluation or other forms of nonfinancial auditing” (p. 58). If these views of program evaluation were accurate at the time, they are certainly not accurate today. They may reflect the authors’ background in, and preference for the terminology of, auditing.

Wheat (1991) says two-thirds of the states currently have “in-house units that regularly produce audit reports that are quite different from traditional financial preaudits and postaudits” (p. 386). He adopts the Greathouse and Funkhouser (1987) definition of performance audit. Performance audit has emerged over the last 20 to 30 years due largely to the emergence of performance budgeting, a changing political culture, an increasingly complex legal environment, and organizational and staffing changes at GAO. Wheat identifies a type of performance auditor he calls an “activist auditor.” Activist auditors independently identify major social and policy problems in their jurisdiction, go beyond the mandate of an audit initiated by an outside authority if such is called for, self-initiate important audits, issue special reports such as budget surveys and reports on financial trends in their jurisdiction, adhere to generally accepted government audit standards, and see the public and its representatives as the audit’s client. Wheat

contends that GAO's yellow book standards, point auditors toward an activist stance. Some obstacles to activist auditing are: elected officials and public managers not eager to share the spotlight, time frames that are much shorter than for other audits, results are seldom, if ever, received with universal acclaim, there are always groups who are threatened by this type of audit, legislative intent is often difficult to determine, data may be suspect, and there is often little agreement on performance measures.

Jonas (1999) argues that most legislative evaluation agencies grew out of the general state government reforms of the 1970s. Legislative program evaluation units were established or existing financial audit units were assigned the performance audit function. Where virtually no legislative program evaluation units existed 30 years ago, a majority of state legislatures now have strong units. Legislative program evaluation adds value to the overall field of program evaluation in the following ways: 1) its focus on utilization affects both the language and techniques of evaluation; 2) systematic development of databases aid information gathering; 3) virtually unlimited access to information aids far-reaching findings; and 4) adaptive research methods result from short timelines mandated by the legislature. Good legislative program evaluation offices improve the prospects for utilization through responsiveness, relevance, language, and quality. Jonas posits that term limits may cause legislatures to be less receptive to the value of legislative program evaluation but he cites no evidence, and conducts no research, to support this supposition.

Berry, Turcotte and Latham (2002) discuss what they see as a change in the paradigm of legislative program evaluation. The older paradigm saw evaluation staff as more independent of the legislature. The new paradigm sees them working more closely with legislators, on shorter time deadlines, with more policy influence and involving

stakeholders more directly. The authors designate two types of state legislative program evaluation units. Type I is traditional legislative auditors and auditor generals who have expanded their duties to include program evaluation. These units are independent but often report to a legislative committee. Type II are units that conduct performance audits and program evaluation but are not part of an auditor general's office. Most of these units report to joint legislative committees or commissions. This type also includes independent legislative offices that are headed by a director. About half the 43 units they identify are type I and half type II. Some of the legislative program evaluation units they identify are administratively housed in the executive branch. They find that "federal requirements appear to have had no significant impact on setting the work of state legislative PE [program evaluation] units" (p. 78). They cite literature that specific evaluation offices don't exist as much in the executive branch as in the legislative because each specific executive branch program manager is expected to act as an evaluator when overseeing the program's budget.

They helpfully conducted a survey of program evaluation units that went beyond just staffing and logistical questions. They list goals of these units derived from their survey. The goals include: achieving legislative responsiveness while maintaining appropriate independence, communicating effectively and making use of technology to communicate results, producing more information in short time periods while maintaining quality, developing hybrid methods of descriptive and analytical work, and educating legislators about agency programs and operations. The authors identify six roles for program evaluation units: 1) educating legislators and citizens, 2) supporting legislative oversight with information, 3) providing timely information, 4) being

proficient and technologically sophisticated evaluators, 5) being clear communicators, and 6) protecting the legislative institution.

Methodology: Surveys

Stutzman and Rainey (1981) address Florida state executive branch officials' views of program evaluation and auditing. They mention that in 1967 in Florida the duty to evaluate public programs was placed on the auditor general, accountable to the legislature. The auditor general mostly concentrates on "program inputs and processes—financial audits and compliance audits" (p. 83) but recently some emphasis on "impact studies" (p. 83). The paper reviews a mail survey of Florida executive branch administrators' views of program evaluation, for example asking what activities they see as program evaluation. Unfortunately, the survey did not address issues such as ways program evaluation could be used to improve programs or possible constraints on the acceptance of program evaluation among executive branch officials.

Botner (1986) reports on a 1985 survey of all 50 states focusing on their post auditing procedures. He defines postaudit as an "audit of the transaction after the expenditure has occurred" (p. 13). The postaudit "conceptually, is a legislative function to be performed by an auditor responsible to the legislature" (p. 13) while preaudits are traditionally an executive function performed by a comptroller. Botner uses Knighton's (1967) definitions of financial, management/operations, program, and performance audits. He surveyed the postauditors in all 50 states as identified by the Council of State Governments. The survey contains information about how postauditors are selected (by

the legislature, elected, appointed by the governor, etc.). Botner finds that 28 state postauditors are selected by legislatures or a legislative body.

Walton and Brown (1990) conducted a survey of state auditors to determine potential conflicts between auditors' roles and legislators' roles. A slight plurality saw no conflict while a sizable minority did see conflicts between the two roles. The two biggest complaints reported were the legislature didn't act on audit findings and the legislature inappropriately attempted to control an audit.

Mohan (1997) conducted a survey of legislative program evaluation units in the states. There were 65 National Legislative Performance Evaluation Society (NLPES) member agencies at the time, 54 responded, 7 of those did not conduct program evaluations or performance audits. About a third of the agencies conducted both performance audits and program evaluations. Approximately one third each performed either program evaluations or performance audits without performing the other. Most units use the GAGAS performance auditing standards while very few use the American Evaluation Association's (AEA) *Guiding Principles for Evaluators* or the Joint Committee's Program Evaluation Standards.

NPLES (2004) conducted a follow-up survey to Mohan's (1997) and NLPES' earlier survey (VanLandingham, Kinton, Boerner, & Zelio, 2000). The results were similar to the one conducted by Mohan, though the 2004 survey was much more comprehensive. The majority of the units surveyed are attached to an auditor's office. About one-third are independent legislative units. The results are similar to 1997 regarding the use of standards; about half follow the GAGAS performance auditing

standards while very few follow the AEA's *Guiding Principles* or the Joint Committee standards. There is some useful information on directives by the overseeing agencies.

Schwartz and Mayne (2005) look at how government organizations in different countries evaluate their evaluations, performance audits and performance reporting information. They cite Schwartz (1999) as a previous report examining or metaevaluating audit reports. Their methodology is basically a survey of these organizations asking them what type of quality assurance systems they have in place. The authors did not look directly at the quality of the work. The surveys show that political concerns are one factor inhibiting quality assurance activities in this area.

Many of these surveys reviewed here fail to address topics that could improve the practice of LPE in the states. Instead they focus on descriptive issues about LPE offices and their staff members. This information is helpful to understand the units. However, an exploration of issues such as the processes LPE units use to evaluate their own work, the methodological techniques they use to determine causality, and the mechanism they have in place to share findings and techniques with other LPE units would do more to advance the practice of legislative program evaluation.

Methodology: Case Studies

Chadwin (1974) edited a volume containing case studies from four states with legislative program evaluation units at the time. The differences in terminology apparent today in legislative program evaluation were apparent then. These states called their reports a program audit (New York), a program analysis (New Jersey), an audit

(Hawai'i), and a comprehensive review (Illinois). All four short case studies concentrate on outlining their objectives and methodologies. They also discuss the impacts of their respective studies and the negative response (in three of the four cases) from officials at the agencies under review.

Roberts and Pollitt (1994) present a case study of a VFM audit completed by the British National Audit Office. The authors view the VFM as a step beyond traditional auditing but still not a "full-blooded evaluation" (p. 527). One way in which it falls short of a program evaluation is that VFMs in Britain are not permitted to "question the appropriateness of government policy objectives themselves" (p. 545). Roberts and Pollitt see three limitations of the VFMs: (1) reports rarely specify methods for improvement; (2) their scope of work is primarily financial; and (3) their dependence on parliament means they are constrained by time.

Bezruki, Meuller and McKim (1999) focus on the experience of the Wisconsin Legislative Audit Bureau and how they see utilization of their evaluative work by their legislature. In Wisconsin, the Joint Legislative Audit Committee oversees the legislative audit bureau. The paper contains three case studies on legislative utilization. The authors state the "evaluations that suggest possible causes but no remedies or that recommend further study have limited use in the legislative process" (p. 21).

McIntire and Glaze (1999) investigated systems modeling, or computer simulation, as a methodological tool for legislative program evaluation. They recount a simulation the authors conducted for the Oregon Legislative Fiscal Office. They concluded that "[b]ecause legislation often affects multiple agencies or issues, models based on budgetary, staffing and process relationships can provide relevant insights for

policy recommendations” (p. 59). However, state legislative program evaluation units need sufficient staffing, funding, training, and time to be able to utilize systems modeling systems in their evaluations.

Methodology: Directly Examining Reports

A methodology directly examining the products of legislative program evaluation—the reports themselves—seems to be crucial to get a good idea of what is actually happening in the field. However, very little of this approach has been undertaken. Funkhouser (2005) cites the scant literature on performance auditing. He sets an agenda for scholarly analysis of performance auditing including: content analysis of actual audit reports, examination of impact of audit organizations, and an examination of the differences in process, products and results in audit organizations that arise from different structures.

Brown and Craft (1980) analyze reports from the small number of states with performance audit sections at the time. They find no difference in reports issued by comprehensive postaudit staff and special purpose committee staff. The criteria they used included report topic, report length, staff effort expended, report focus, data gathering techniques, presence of recommendations, and executive agency response. They found that most reports confirm, and provide documentation for, what legislators already believed but did not know for certain. They posited reasons why some reports are implemented while others are not. Reports that were released at the beginning of a legislative session, received media coverage, and included concrete findings and

recommendations were likely to be acted upon. The methodology used in a report had no affect on whether the report's recommendations were implemented.

Walker (1985) reviewed a broad sample of reports from the GAO (35 in their entirety and 100 summaries). Walker did not present his criteria for reviewing these reports. His review of reports was one aspect of a larger review of the GAO. That review included "213 anonymous interviews with executives and auditors at all levels of the [GAO], congressmen and staffers, agency officials, and newsmen" (p. 360).

Brown (1988) provides some numbers on state legislative evaluation within the Illinois Auditor General's office. Though Brown's observations seem to be based more on experience than a strict review methodology, she reports that 30 percent of the problems or issues highlighted in audit reports concern matters of policy for the Illinois legislature to address. These are either technical or substantive changes to law or issues of public policy. When recommendations are offered in the audit reports they are in the form of alternatives to be considered by the legislature. The other 70 percent of problems or issues addressed in audit reports concern administrative matters that are directed to the executive branch agency under review.

Schwartz (1999) examined "effectiveness audits or evaluations" (p. 513) in different countries to see how they really address effectiveness in light of the political difficulties of doing so. Schwartz's methodology was to examine the content of three audit reports from each of six countries. Curiously, instead of selecting random reports he asked certain SAI representatives from these countries for their three best audits. He tried to determine if they are really looking at effectiveness and how they negotiate the potential political problems with really looking at effectiveness. He found that "most

state auditors have not met the challenge non-politicized, professional and objective outcome evaluations in place of agency-based evaluations which are perceived to be infected by political interests” (p. 522). He concludes that the main reason for this is the auditors’ desire to avoid the political dangers inherent in questioning the effectiveness of programs.

Metaevaluation

Michael Scriven first introduced the term metaevaluation in 1969. At the time the term applied to the evaluation of a plan to evaluate educational products. Since, Scriven has used metaevaluation to refer to an evaluation of evaluations (Scriven, 1991).

As in conducting an evaluation, when conducting a metaevaluation one uses standards (or values and criteria) to judge the evaluations under review. There are several sets of standards that can be used to metaevaluate evaluations—professional standards like the American Evaluation Association’s *Guiding Principles for Evaluators*, the Generally Accepted Government Auditing Standards (GAGAS) or the Joint Committee’s Program Evaluation Standards (PES). Evaluations can also be metaevaluated against various evaluation models—Stufflebeam’s CIPP model (Stufflebeam, 2000a), Patton’s Utilization-Focused Evaluation Model (Patton, 1997), Guba and Lincoln’s Constructivist Evaluation model (Guba & Lincoln, 1989) and others.

Some metaevaluators conduct metaevaluations with access to a broad scope of information—interviews with the evaluators and stakeholders and review of the evaluators’ data. Others have completed metaevaluations using only the final evaluation

report, or even an executive summary of the report. For two main reasons the metaevaluations for this study were conducted using only the final evaluation report. First, to conduct a comprehensive metaevaluation (interviewing evaluators and stakeholders, reviewing raw data, etc.) would be entirely too time consuming given the large sample of reports reviewed. Other endeavors to metaevaluate large numbers of evaluations have used solely the evaluation reports. (Bickman, 1997; Bollen, Paxon, & Morishima, 2005; Cooksy & Caracelli, 2005; Forss & Carlsson, 1997; Hoogerwerf, 1992; Kruse, 2005; Larson, et al, 1979; Scott-Little, Hamann, & Jurs, 2002) Second, LPE reports are somewhat different from other evaluation reports. These reports are produced for the use of legislators and as such are typically briefer and less technical than many other evaluation reports. There is value in developing a process to metaevaluate these reports solely by reviewing the report itself. Users in contexts other than the specific state legislature for which the original report is produced need a way to evaluate the quality of reports. Such a process could aid staff and legislators in other jurisdictions in determining how useful an evaluation report is to their own policy development.

The few published papers that metaevaluate multiple studies in a particular content area or general field vary greatly in their respective scopes and methodologies. Table 1 lists each study's author(s), year published, number of evaluation reports metaevaluated, whether the reports metaevaluated consisted of a random sample of a larger population of reports, whether the metaevaluators relied solely on the evaluation reports, and other data sources examined if the metaevaluator did not rely solely on the evaluation report. The number of reports metaevaluated ranges from just a few to over 200. Only one study (Hoogerwerf, 1992) examined a random sample of a larger

population of evaluation reports. Similarly, only one study (Chatterji, 2005) examined data beyond the evaluation report while conducting the metaevaluations.

Table 1. Metaevaluation Literature: Multiple Studies.

| Author(s) | Year | Number of reports | Random sample used | Relied solely on reports | Other data examined |
|------------------------------|------|-------------------|--------------------|--------------------------|---|
| Bollen, Paxon, & Morishima | 2005 | 25 | No | Yes | None |
| Chatterji | 2005 | 4 | No | No | Interviewed evaluator, reviewed documents |
| Cooksy & Caracelli | 2005 | 87 | No | Yes | None |
| Leeuw & Cooksy | 2005 | 3 | No | Yes | None |
| Kruse | 2005 | approx. 200 | No | Yes | None |
| Scott-Little, Hamann, & Jurs | 2002 | 23 | No | Yes | None |
| Bickman | 1997 | 56 | No | Yes | None |
| Forss & Carlsson | 1997 | 277 | No | Yes | None |
| Hoogerwerf | 1992 | 15 | Yes | Yes | None |
| Larson, et al. | 1979 | 200 | No | Yes | None |

Note. Approx. = approximately.

Following is a brief discussion of the published literature that uses metaevaluation across multiple evaluations, focusing on the content area, the standards used, and the number of evaluations included.

Leeuw and Cooksy (2005) contend that metaevaluation “is a method that should be used to ensure that the meta-studies (either narrative reviews or meta-analyses) are based on defensible evaluations” (p 96). This is especially important given the context of state legislative program evaluation. The various state LPE units often seek information from each other when addressing a certain policy area. Metaevaluation can increase the

confidence LPE units have in the quality of an evaluation report that they decide to reference in their own reports.

Leeuw and Cooksy also identify three main uses for metaevaluation: aiding stakeholder decisions about evaluation use, aiding evaluator/researcher decisions about including findings from specific evaluations in an evaluation synthesis, and “identifying strengths and weaknesses in evaluation practice” (p. 97). Metaevaluation is used in the current study for the latter two uses with an emphasis on the third.

Leeuw and Cooksy highlight a categorization developed and used by Hoogerwerf (1992), Leeuw (1998), Vedung (1997), Van de Vall (1980) and others. They distinguish only three criteria: methodology, strategy (type and relevance of evaluative knowledge produced), and utility. This categorization allows metaevaluators to tailor “standards within each of the three dimensions of quality” (p. 100).

They then use these three dimensions of quality to metaevaluate three international development evaluations. They identified five criteria on the methodological dimension (purpose of the evaluation; criteria of success; data sources; time period; and types of analyses), three criteria on the strategic dimension (relevance; type of knowledge produced; attribution addressed; and findings related to the evaluation’s goals) and three criteria on the utility dimension (what works, why, and when; recommendations; and audience) for a total of 11 criteria for the metaevaluation.

Hoogerwerf (1992) looks at how policy evaluation developed in the central government in the Netherlands. More specifically he addresses how “political-administrative authorities judge policy evaluation and its use in the policy process” and how “researchers judge policy evaluations and their use in the policy process” (p. 215).

Hoogerwerf presents two diagrams, or lists of metaevaluation criteria. One diagram is “scientific requirements of an evaluation report” the second is “requirements from policy practice concerning an evaluation report” (pp. 220 and 221).

The scientific requirements concern theoretical requirements (such as formulation of the problem, definition of concepts), methodological requirements (such as the validity and reliability of the research results) and technical requirements (such as the operationalization of the evaluation criteria). Hoogerwerf designed this list for reports aimed at the effects of policy. He writes it should not be used when evaluating the contents or processes of a policy.

The requirements from policy practice are “based on the principle that evaluation research . . . should be useful to policy practice” (p. 222). An evaluation report should “provide information which is important to an effective, efficient and legitimate approach to a certain policy problem” (p. 222). Factors affecting the use of research in the policy process include “the adaptation of the research report to the need for information in policy practice, the nature of the policy, and readability for policy makers” (p. 222).

Hoogerwerf discusses his application of these two lists of requirements to a random sample of 15 “evaluation reports on elements of the sports policy of the Ministry of Welfare, Public Health and Culture in the Netherlands”. He reports that the “main results of this meta-evaluation are not very positive” (p. 222).

Hoogerwerf concludes by developing “requirements from policy practice to be imposed on the evaluation process” and “requirements from policy practice to be imposed on evaluation structure” (p. 224). The requirements for structure are particularly relevant to the LPE units in the states. These include the requirement that any

governmental organization or network should include a unit “which is charged with evaluating the implemented or intended policy, and with stimulating its feedback.” This unit should “have sufficient knowledge on scientific evaluation research,” it should “be so independent that it can perform its duties in an adequate way,” and it should be “responsible for the coordination of the evaluation and of the feedback” (p. 225).

Kruse (2005) discusses a metaevaluation of about 200 reports. This covers all reports submitted to the Norwegian Agency for Development Cooperation (NORAD) over two years. Kruse’s metaevaluation criteria focused on the evaluation process, scope, and team and on methodological issues such as data sources, collection methods and types of analysis. Under the category of ‘types of analysis carried out’ the criteria included: discussion of limitations in methods, cost-effectiveness analysis, lessons learned and policy/strategy discussions. Reports were graded ‘excellent,’ ‘good,’ ‘adequate,’ or ‘inadequate’ using the quality criteria developed by Forss and Carlsson (1997). Kruse found that 45 percent of the evaluations were “good” and 30 percent “adequate.”

Forss and Carlsson (1997) examined evaluation reports commissioned by development aid agencies in Sweden. They used the four main standards from the PES (utility, feasibility, propriety, and accuracy) as a starting point to judge the quality of these evaluation reports. They describe three paths they could have followed for their research. First, they could have used a case-study approach to determine “what motivation stakeholders had, how they were engaged in the process, which methods were used, how data were analysed and by whom, and what difference the evaluation process made in the management of activities” (p. 483). This approach would have provided a

thorough understanding but would have taken much time and resources and they would have been unable to ensure that the case-studies were representative of the larger population of evaluations. Second, they could have interviewed aid organization personnel to get their views on evaluation quality. However, this approach would have discovered the views of just one of many aid program stakeholders. They decided on a third approach—to “[a]ssess the evaluation reports through the application of a standardized format, analysing their coverage, composition of teams, methods of work etc.” (p. 484). They correctly defend their use of this approach based on the fact that it allowed them to cover a large number of evaluations at relatively low cost.

They concluded that this method can “be expected to provide comprehensive information on the accuracy and propriety standards, but less concerning utility and feasibility” (p. 484). They used one of four ratings for each variable: ‘excellent,’ ‘good,’ ‘adequate,’ or ‘inadequate.’ The study included 277 evaluation reports commissioned by development aid agencies in Sweden. The authors are not clear on whether this is a sample (random or otherwise) or the total population of reports over a certain time period. Ratings were assigned, primarily, by several research assistants who received training on the criteria and ratings. The research assistants were able to consult with each other and also to seek input from the authors.

Generally, the authors find that the quality of the evaluations is not good. Also, it seems that over time (the evaluations examined spanned about 20 years) the quality has not improved. However, they posit that this may be because they only were able to locate older evaluations that were good, the worse ones being forgotten.

They insightfully point out two major drawbacks to their approach to meta-evaluation. First, it is difficult to train researchers well to evaluate the evaluation reports. Second, the process can be rather expensive since it typically takes two to four hours to review a single evaluation report.

Scott-Little, Hamann, and Jurs (2002) looked at evaluations of after-school programs. They metaevaluated the reports based on the Joint Committee's PES (1994). They also examined each evaluation's research design. Their research questions included: "What are characteristics of after-school evaluation studies that have been disseminated publicly?" and "To what extent do these evaluation studies conform to evaluation standards and common requirements for research design?" (p. 389). The authors identified 43 evaluations of after-school programs and narrowed that number down to 23 to eliminate certain kinds of after-school programs (those that were drop-in or special programs).

Bollen, Paxon, and Morishima (2005) reviewed 25 of the 240 evaluation reports from USAID's Democracy and Governance (DG) projects. Their population of 240 reports reflected all DG evaluation reports from the past 30 years. They selected a non-random sample, including some reports from each region, each institutional author, and each subject area plus additional reports that USAID particularly wanted included in the sample.

Their "review focused on three general and necessary ingredients to quality evaluations: (a) information on inputs, (b) information on results, and (c) controls for confounding factors" (p.190). The authors reveal that their

. . . fundamental principle is that a quality evaluation collects and reports sufficient information on the first two ingredients and addresses the threats

to any conclusion of impact presented by the third. For the purposes of this review, inputs were USAID activities and the resources used to implement them, and results were the intended outputs, outcomes, and impacts of the activities. (p. 190)

The review considered four categories of possible confounding factors: alternative explanations, the direction of the causal relationship, sampling issues, and measurement issues. For each report, each confounding factor was coded yes, no or slightly for the following categories: alternative explanations, other donors recorded, prior change considered, controls considered, and concepts defined. The authors coded yes/no four categories concerning measurement: measures defined, alternative measures, validity discussed, and reliability discussed.

They found that “hardly any evaluations” discussed alternative ways to measure the same criteria or the validity/reliability of the measures used. They concluded that “[o]verall, the treatment of measurement in these evaluations is largely absent” (p. 198). Most of the coding was fairly objective. Coders recorded activities, funding, personnel, the level of outcomes (individual, group, institutional, or national), and the time frame during which observations were made. They used multiple reviewers claim high inter-rater reliability, though they undermine their contention by not reporting specific numbers on this issue.

Chatterji (2005) applies the PES to an international aid evaluation in Bangladesh. She selected 4 evaluation reports to metaevaluate based on four criteria, the second being “[t]here is enough available information, through documentary or other sources, to allow a systematic appraisal of the evaluation practices with respect to a majority of the 30 standards, if not all. (Thus, if the available data were scant, and it led to too many *cannot determine* or *not applicable* conclusions during analysis, the case was dropped.)” (p.

2382). In addition to reviewing the evaluation reports, she interviewed the lead evaluator to follow-up on questions. She concludes that the accuracy standard could be metaevaluated from the report but the feasibility, propriety and utility standards needed the information provided in the interview.

Bickman (1997) examined all 56 “impact or effectiveness evaluations” published in *Evaluation Review* and *Evaluation and Program Planning* over a ten year period from the mid-1980s to the mid-1990s. (p. 1) He and an assistant evaluated each study based on several variables, drawn from Lipsey’s (1988) meta-analysis procedures. Bickman created two composite scores. One, for several measures of methodological and program theory quality, ranged from 12 to 50. The second, a composite score derived from “the sum of quality ratings given to various components of design” ranged from 8 to 24 (p. 4). Bickman, unfortunately, does not further describe the rating process or the criteria. He concludes that the reports are not “approaching a ceiling on quality” and “[q]uality has not changed in the last decade.” (p. 4)

Larson, et al. (1979) issue an interim report that examined a sample of 200 evaluation reports out of approximately 1,500 evaluations performed on criminal justice issues from the late 1960s through the late 1970s. The authors note that “the final report of an evaluation is unlikely to capture most of the information about an evaluation necessary to evaluate it comprehensively” (p. 2). They conclude that “the documents produced in an evaluation constitute only one output of an evaluation: they often tell little of the inputs to an evaluation, the process by which it was undertaken, or the ultimate impact of the evaluation” (p. 2). The LPE reports examined here include much information about the inputs of the evaluation and the process by which it was

undertaken. This is probably because the GAGAS standards require this information. The sample they took was “a structured rather than a random sample” (p. 5) because they selected a certain amount of reports on different types of programs.

Three graduate student readers each read about a third of the 200 evaluation reports taking about 4 hours per report and completing a 31 item checklist on each report. The authors note that “this amount of time is small to evaluate any evaluation satisfactorily” but “it is probably more time than most decision makers would spend on the report.” (p. 15) The “goal of the readings was not to evaluate each evaluation, but rather to obtain information regarding evaluation input, processes, and outcome, and to assess in a general way the relevance of the methodology employed, and the quality of the documentation.” (p. 15)

Only four of the 200 evaluation reports indicated the total funding for the evaluation. The authors rightly believe that this information should be included in evaluation reports because “[o]ne should be able to assess the value of an evaluation given known budgetary constraints.” (p. 18) It might be added that knowing the costs of an evaluation is helpful, if not necessary, to determine the evaluation’s merit and worth. In addition 31 percent of the reports documented the time spent on the evaluation.

Cooksy and Caracelli (2005) state that “[t]he second major purpose for subjecting a set of studies to metaevaluation is to identify strengths and weaknesses in evaluation practice in order to develop evaluation capacity” (p. 32). The first major purpose concerns informing researchers of which studies to include in evaluation syntheses. They cite Lipsey, Crosse, Dunkle, Pollard, & Stobart (1985) and Bickman (1997) as doing

large metaevaluations and finding that the studies were marginally methodologically and conceptually sound.

Cooksy and Caracelli metaevaluated 87 evaluation reports (using just the reports to conduct the metaevaluations) concerning international agricultural programs. They note that critical information regarding methodological issues was often missing. This lack of transparency made it difficult to judge the overall methodological quality.

There is a small body of literature concerning the theoretical aspects of metaevaluation. This literature addresses questions such as how to conduct a metaevaluation and when metaevaluation is needed. Below is a discussion focusing on the question of how much information one needs to conduct a metaevaluation. Does one need as much information as the original evaluators needed? Does a metaevaluator need to basically redo the evaluation or can she conduct an appropriate metaevaluation using only the completed evaluation report?

Sanders (1999), while not addressing this question explicitly, conducted a metaevaluation of a study published in the *American Journal of Evaluation*. The metaevaluation was conducted on the basis of a 20 page summary (printed in AJE) of a 200 page evaluation report. Sanders' metaevaluation was sent to the original evaluation's authors to correct any inaccuracies. Sanders concludes that "[i]n the absence of the full evaluation report it is incumbent on a metaevaluator to seek feedback on the accuracy of a metaevaluation from the primary evaluator" (p. 582). This leads to the conclusion that given the full evaluation report Sanders believes a legitimate metaevaluation can be conducted from examining the evaluation report alone.

Sanders' approach is similar to the approach, used in this study, of conducting a metaevaluation from the written evaluation report itself, without collecting data from evaluation subjects as the more comprehensive Stufflebeam (2001a) approach would have metaevaluation done. Sanders uses the Program Evaluation Standards (PES) and the AEA's *Guiding Principles for Evaluators*.

Following the PES, Sanders addresses utility, feasibility, propriety, and accuracy. Under the heading of utility he is concerned with issues such as were the intended users involved in developing evaluation questions, were stakeholders involved in identifying the criteria on which the evaluand is being judged, were results "disseminated in audience-friendly directed communications" (p. 578) and were the evaluators well qualified and independent from the evaluand.

For feasibility he addresses the "practicality, workability and cost of the evaluation" (p. 579). This includes whether the procedures to collect data were workable, not offensive or disruptive, and acceptable to the participants.

The propriety standards are "intended to protect participants in the evaluation and those affected by the study's results" (p. 579). This would include the question of whether there was a bias, conflict of interest, or misuse of funds by the evaluators. Also, did the benefit from the evaluation worth the cost of conducting the evaluation.

Finally, the accuracy standards are "concerned with the technical merits of the evaluation" (p. 580). This area includes whether there was information given in the evaluation report about the actual means of delivery of the evaluand (how the evaluand works), were there gaps in the analysis (did the evaluators merely report outputs but not

outcomes or results), did the authors describe their quality control procedures for data collection and their evaluation methodology, and was there “underanalysis” in places.

Cook and Gruder (1978) identify three research traditions relevant to metaevaluation. The first is a re-analysis of evaluation data, the second is rating evaluations to see how “technically competent they were in general” (p. 6), and the third is research on research. This dissertation focuses on the second tradition. They talk about the high standards of past metaevaluations and whether the evaluations judged needed to live up to these high standards. They also address the issue of reaching plausible versus definitive causal conclusions. This point fits in well with the LPE aims of informing public policy making. LPE reports can greatly add to legislative decision-making, and thus public policy, by presenting plausible conclusions even if they cannot reach definitive conclusions.

Cook and Gruber discuss their contention that metaevaluators have a bias toward the negative. Metaevaluators “want” to make critical findings or surprising findings. They discuss seven metaevaluation models (p. 17), four are subsequent to the primary evaluation and three are simultaneous to the primary evaluation. The most basic of the subsequent metaevaluation models they call the essay review of an evaluation report. An essay review is completed subsequent to the original evaluation, there is no manipulation of data from the primary evaluation, and there is a single data set (the single evaluation report). Their view of an essay review does not include using standards and assessing the evaluation report against these standards. It is more critiquing various aspects of the evaluation report and its conclusions.

Larson and Berliner (1983) discuss what they call the evaluation of evaluations (or EOE). They contend that the “evaluative study of evaluations requires analysis of evaluation inputs, process, and outcomes” (p. 155). They define evaluation inputs as an “inventory of resources and methodologies brought to bear on the evaluation and the basic elements of the evaluation/program setting” (p. 155). They list 11 aspects of evaluation inputs including budget, duration, timing, evaluation personnel, methodology and design, data and data limitations, and expected policy implications. They define evaluation process as “[a]ctual conduct of the evaluation compared with that planned in the evaluation design” (p. 157). Aspects of the evaluation process include interactions between evaluators and program staff members, the ease with which one can adapt the evaluation design, methodology, and communication of findings. Evaluation outcomes are the “decisions influenced by the evaluation” (p. 160). These decisions included a funder’s decision to increase, decrease, or suspend funding, decisions by program staff to modify the program, and decisions by clients to alter their participation in the program.

Larson and Berliner point out that the level of documentation included in a final report of an evaluation is important to determining if an external EOE can be conducted. They discuss previous research (Larson, et al., 1979) that shows documentation on a number of variables lacking in a sample of criminal justice program evaluations. They also found a lack of reporting on an evaluation’s methodology. “Such lack of self-reporting reduces the ability of decision makers to assess the quality of information produced by the evaluation” (p. 161).

Pawson (2002a, 2002b) contends that evaluators generally agree that there is no single best way for evaluating social programs or for conducting “systematic review[s]”

of evaluations. His goal is to lay out a new model for informing evidence-based practice, what he calls “realist synthesis.” Realist synthesis contends it is “much more sensible to base any systematic review on ‘families of mechanisms’ rather than on ‘families of programmes’” (Pawson, 2002b, p. 344). These mechanisms for influencing behavior are more important to study together than are the programs that make use of the mechanisms.

Pawson offers an insightful quote that captures how legislative program evaluation, and the systematic evaluation of LPE reports, can aid the development and improvement of public policy:

Few major public initiatives these days are mounted without a sustained attempt to evaluate them. Rival policy ideas are thus run through endless trials with plenty of error and it is often difficult to know which of them really have withstood the test of time. It is arguable, therefore, that the prime function of evaluation research should be to take the longer view. By building a systematic evidence base that captures the ebb and flow of programme ideas we might be able to adjudicate between contending policy claims and so capture a progressive understanding of ‘what works’. (Pawson, 2002a, p 157)

Background and Comparison of Evaluation and Performance Audit Standards

This dissertation uses two explicit sets of professional standards—one for performance auditing, the Generally Accepted Government Auditing Standards for Performance Auditing, and one for program evaluation, the Program Evaluation Standards. In addition, another framework considered, the Key Evaluation Checklist, comes from the evaluation field. The two sets of standards were developed through a similar process consisting of collaboration and committees. Each has been updated since its first inception. This section will give a background on the development of the GAGAS

for performance auditing and the Joint Committee's Program Evaluation Standards showing similarities and differences in their creation and maintenance. The section will also discuss the similarities and differences of these two standards.

Generally Accepted Government Auditing Standards for Performance Auditing

The GAGAS are described as “the floor of acceptable behavior” for professionals. They are “broad statements of auditors’ responsibilities” that “provide an overall framework for ensuring that auditors have the competence, integrity, objectivity, and independence in planning, conducting, and reporting on their work” (General Accounting Office, 2003, p. 1). The 2003 version was the fourth version since the standards were issued in 1972 (previous revisions were issued in 1988 and 1994). A 2007 revision was issued in January (and updated in July) with an effective date of January 1, 2008.

The GAGAS indicated that other professional standards, specifically including the American Evaluation Association's *Guiding Principles for Evaluators* and the Joint Committee's Program Evaluation Standards, are “not incorporated into GAGAS but can be used in conjunction with GAGAS” (p. 9).

The GAGAS have been applicable to performance auditing since they were first issued in 1972. Though that first edition did not refer specifically to performance audits, it did “provide for a scope of audit that includes not only financial and compliance auditing but also auditing for economy, efficiency, and achievement of desired results” (Comptroller General of the United States, 1972, p. 2). The 1981 revision notes that the term performance audit is one of several terms “used for economy and efficiency audits

and program results audits” (Comptroller General, 1981, p. 68). However, that revision also indicates that the terms “economy and efficiency” and “program results” audits should be used “in audit reports to avoid misunderstanding” (p. 12).

It was the 1988 revision that introduced sections on field work and reporting standards specifically for performance audits. This revision also offered a definition of performance audits that simply included economy and efficiency and program audits (Comptroller General, 1988, p. 2-3). As recently as that year state legislative program evaluation units did not widely accept the GAGAS performance auditing standards and did not follow them in their performance auditing and evaluation work (Greathouse and Funkhouser, 1988).

The 2003 revision continued the previous revision’s model of sections for field work and reporting standards for performance audits. This revision also offered a more detailed definition of performance audits. They are audits that

entail an objective and systematic examination of evidence to provide an independent assessment of the performance and management of a program against objective criteria as well as assessments that provide a prospective focus or that synthesize information on best practices or crosscutting issues. (General Accounting Office, 2003, p. 21)

Under the general heading of performance audits the 2003 revision of the GAGAS addresses “program effectiveness and results audits,” “internal control audits,” “compliance audits,” and other activities audit organizations undertake that provide a “prospective focus or may provide guidance, best practice information, and information that cuts across program or organizational lines, or summary information on issues already studied or under study by an audit organization” (p. 24).

The GAGAS details performance auditing standards in an introduction and three chapters. The GAGAS also provides standards for financial audits, attestation engagements, and nonaudit services provided by audit organizations. The introduction and the first chapter provide general standards for all types of audits. The introduction includes sections on purpose, applicability, relationship between GAGAS and other professional standards, accountability, and roles and responsibilities. The first chapter is divided into categories for independence, professional judgment, competence, and quality control and assurance. The next two chapters (field work standards and reporting standards) provide specific standards for performance audits. The field work standards are divided into planning, supervision, evidence, and documentation. The reporting standards are divided into form, content, quality elements, and issuance and distribution.

Absent from the GAGAS is any mention of an auditor's responsibility to look beyond program or policy goals and objectives. There is also no mention of examining comparison programs or policies when conducting a performance audit. The only brief mention of comparisons in the GAGAS comes in a definition of analytical evidence.

Combining the introduction and all three chapters we are left with 17 separate areas of standards. This compares with 30 separate standards for the Program Evaluation Standards (seven utility standards, three feasibility standards, eight propriety standards, and twelve accuracy standards).

Program Evaluation Standards

The forerunner of the Program Evaluation Standards was the *Standards for Evaluations of Educational Programs, Projects, and Materials* published by the Joint Committee on Standards for Educational Evaluation in 1981. The Joint Committee was originally formed in 1975 after a committee convened by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education to examine testing decided to spin off a subcommittee regarding program evaluation (Yarbrough, Shulha, & Caruthers, 2004). The Program Evaluation Standards' birth through the cooperation of professional organizations contrasts with that of the GAGAS, which were originated by a government agency (though with the involvement of professional organizations).

This original conception of the PES was divided into the same four categories as the current PES (utility, feasibility, propriety, accuracy). The original GAGAS were divided into four sections: an introduction and general standards (as in the 2003 version), examination and evaluation standards, and reporting standards.

Table 2 shows each PES standard and its corresponding GAGAS standard. Not all PES standards have corresponding GAGAS standards. Most notably, the PES propriety standards do not have a corresponding category of standards in the GAGAS. The PES utility and accuracy standards correspond roughly with the GAGAS field work and reporting standards.

Table 2. Program Evaluation Standards Versus GAGAS.

| PES standard | Corresponding GAGAS standard(s) |
|--|---|
| U1 Stakeholder identification | No corresponding standard |
| U2 Evaluator credibility | General standard 3.39 – Competence |
| U3 Information scope and selection | Reporting standards for performance audits, report quality elements, 8.41 – Complete |
| U4 Values identification | No corresponding standard |
| U5 Report clarity | Reporting standards for performance audits, report quality elements, 8.50 – Clear |
| U6 Report timeliness and dissemination | Reporting standards for performance audits, report quality elements, 8.39 – Timeliness |
| U7 Evaluation impact | Reporting standards for performance audits, report quality elements, 8.49 – Convincing |
| F1 Practical procedures | No corresponding standard |
| F2 Political viability | No corresponding standard |
| F3 Cost effectiveness | No corresponding standard |
| P1 Service orientation | No corresponding standard |
| P2 Formal agreements | No corresponding standard |
| P3 Rights of human subjects | No corresponding standard |
| P4 Human interactions | No corresponding standard |
| P5 Complete and fair assessment | Reporting standards for performance audits, report quality elements, 8.41 – Complete |
| P6 Disclosure of findings | Introduction, roles and responsibilities, 1.25 – Auditors’ responsibilities |
| P7 Conflict of interest | Introduction, roles and responsibilities, 1.24 – Auditors’ responsibilities |
| P8 Fiscal responsibility | General standards, 3.19a – External impairments General standards, 3.51 – Quality control and assurance Reporting standards for performance audits, report quality elements, 8.29 – Recommendations |
| A1 Program documentation | Reporting standards for performance audits, report quality elements, 8.43 – Accurate |
| A2 Context analysis | Reporting standards for performance audits, report quality elements, 8.13 – Findings |
| A3 Described purposes and procedures | Reporting standards for performance audits, report quality elements, 8.08 – Objectives, scope, and methodology |
| A4 Defensible information sources | Reporting standards for performance audits, report quality elements, 8.38 – Report quality elements |

Table 2—continued

| PES standard | Corresponding GAGAS standard(s) |
|---|---|
| A5 Valid information | Field work standards for performance audits, 7.31 – Identifying sources of audit evidence Field work standards for performance audits, 7.52a – Tests of evidence Reporting standards for performance audits, report quality elements, 8.49 – Convincing |
| A6 Reliable information | Field work standards for performance audits, 7.31 – Identifying sources of audit evidence |
| A7 Systematic information | General standards, 3.49 – Quality control and assurance Reporting standards for performance audits, report quality elements, 8.45 – Accurate |
| A8 Analysis of quantitative information | No corresponding standard |
| A9 Analysis of qualitative information | No corresponding standard |
| A10 Justified conclusions | Field work standards for performance audits, 7.48 - Evidence Reporting standards for performance audits, report quality Elements, 8.49 – Convincing |
| A11 Impartial reporting | Reporting standards for performance audits, report quality elements, 8.46 – Objective General standards, 3.07 – Personal impairments |
| A12 Metaevaluation | Introduction, roles and responsibilities, 1.27 – Audit organizations’ responsibilities General standards, 3.49 – Quality control and assurance |

For ease of presentation this table, and the subsequent discussion, follows the PES through the utility, feasibility, propriety, and accuracy standards and ties each to corresponding GAGAS standards.

Stakeholder Identification (U1)

The PES stakeholder identification standard requires “[p]ersons involved in or affected by the evaluation should be identified, so that their needs can be addressed”

(Joint Committee, 1994, p. 25). The PES locates this standard within the utility category, stressing that an “evaluation planned with stakeholders and conducted to meet their information needs is likely to attend to and receive a positive response” (Joint Committee, 1994, p. 25).

This is one area where the GAGAS is largely silent. The GAGAS repeatedly refers to subject of a performance audit as the “audited entity.” The GAGAS does recognize the link between certain stakeholders and performance audit use. There are references to the auditor’s responsibility to assist the audited entity in understanding the findings and recommendations of the audit (General Accounting Office, p. 163). However, there are no requirements that an auditor identify and assess the needs of either the audited entity or other stakeholders.

The PES stakeholder identification standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) is incorporated into the metaevaluation criteria under the same name in the preliminaries section.

U2 Evaluator Credibility

Credibility is linked to utility because an evaluator with credibility is important to the acceptance and use of the evaluation findings. As the PES phrases the issue, evaluators “should be both trustworthy and competent to perform the evaluation, so that the findings achieve maximum credibility and acceptance” (Joint Committee, 1994, p. 31). The PES also suggests that evaluators establish credibility with stakeholders at the

beginning of the process and consider not proceeding if an appropriate degree of credibility cannot be secured.

The GAGAS competence standard is most closely associated with the PES credibility standard. The competence standard, included in the general GAGAS standards, provides that the audit staff “should collectively possess adequate professional competence for the tasks required” (General Accounting Office, p. 52). Technical knowledge and competence and continuing professional education are the two substandards under competence. The GAGAS places the responsibility for ensuring auditor competence on the individual audit organizations. This highlights a recurring difference between the GAGAS and the PES. GAGAS is much more directed at organizations while the PES is directed towards individual evaluators and individual evaluation reports. An additional difference between this GAGAS standard and the PES credibility standard is that the competence standard is not explicitly linked to the acceptance or the utility of the audit.

Both the PES evaluator credibility standard, which is not identified by the Joint Committee as particularly relevant for reporting an evaluation, and the GAGAS competence standard are not incorporated into the metaevaluation criteria. Determining the degree of competence and credibility of auditors and evaluators would have been exceedingly difficult using just the evaluation and audit reports.

U3 Information Scope and Selection

The PES provides that “[i]nformation collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.” (Joint Committee, p. 37)

A similar standard under the GAGAS is listed under the heading complete in the reporting standards for performance audits. Complete is a substandard under report quality elements. The GAGS requires audit reports to “contain all evidence needed to satisfy the audit objectives and promote an adequate and correct understanding of the matters reported” (General Accounting Office, p. 174).

Here we see the GAGAS standards refer to satisfying the audit objectives while the PES stresses the need for the information to respond to the needs of clients and stakeholders. Under the GAGAS performance auditing objectives can be broad or narrow including such areas as “assessing program effectiveness and results; economy and efficiency; internal control; compliance with legal or other requirements; and objectives related to providing prospective analyses, guidance, or summary information” (General Accounting Office, 2003, p. 21). In contrast, the PES is more focused on the “program in terms of all important variables . . . whether or not stakeholders specifically ask for such information” (Joint Committee, 1994, p. 37).

The PES information scope and selection standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS complete standard are incorporated into the metaevaluation criteria under the complete criteria in the methodology section.

U4 Values Identification

Values, which literally supply the root for evaluation, are heavily stressed in the PES. The values identification standard provides that the “perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear” (Joint Committee, 1994, p. 43). Evaluators, clients, and other stakeholders should “thoughtfully determine the approach to be followed in assigning values to the obtained information, and they should reveal and justify their chosen approach” (Joint Committee, 1994, p. 43).

The GAGAS standards do not address the issue of values, instead focusing on objectives—defined as “questions about the program that auditors seek to answer” (General Accounting Office, 2003, p. 126). Though these questions often raise value-laden issues such as effectiveness, results and efficiency (see discussion above regarding performance auditing objectives) the language of values is not broached in the GAGAS.

The PES values identification standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) is incorporated into the metaevaluation criteria under identification of values in the preliminaries section.

U5 Report Clarity

Evaluation reports should clearly describe the “program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation” (Joint Committee, 1994, p. 49). This clarity is important for both credibility and utility.

The GAGAS standards include a nearly identical requirement. One of the report quality elements substandards, 8.50, is clear. This standard is somewhat narrower in its application than is the corresponding PES standard. The GAGAS clear standard requires a “report be easy to read and understand” and be “prepared in language as clear and simple as the subject permits” (General Accounting Office, 2003, p. 176). This standard, unlike the comparable PES standard, does not address specific areas that need to be clearly reported (such as the program, context, purposes, and procedures).

The PES report clarity standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS clear standard are incorporated into the metaevaluation criteria under clear in the results section.

U6 Report Timeliness and Dissemination

“Evaluators should communicate evaluation findings to intended users at times when the information can best be used” (Joint Committee, 1994, p. 53). Under report quality elements, the GAGAS timely substandard (8.39) similarly requires an audit report “to provide relevant information in time to respond to officials of the audited entity, legislative officials, and other users’ legitimate needs” (General Accounting Office, 2003,

p. 173). Both standards make provisions for reporting interim information (i.e. before the issuance of a final report). Both also suggest planning and agreements with clients and other users.

The PES report timeliness and dissemination standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS timely standard are incorporated into the metaevaluation criteria under timeliness in the results section.

U7 Evaluation Impact

To increase the likelihood an evaluation will be used “[e]valuations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders” (Joint Committee, 1994, p. 59). This is an affirmative requirement for an evaluator that can extend beyond the issuance of the final report. Evaluators should not “assume improvements will occur automatically once the evaluation report is completed” but should actively assist stakeholders in pursuing those improvements.

While the GAGAS has no comparable standard, auditors are required to produce reports that are convincing enough for the “report users to recognize the validity of the findings, the reasonableness of the conclusions, and the benefit of implementing the recommendations” (General Accounting Office, 2003, p. 176).

The PES evaluation impact standard (though identified by the Joint Committee as particularly relevant for reporting an evaluation) is not incorporated into the

metaevaluation criteria because it would not be possible to determine report impact solely from examining the report itself.

F1 Practical Procedures

Evaluation procedures should “keep disruption to a minimum while needed information is obtained” (Joint Committee, 1994, p. 65). Once again, the GAGAS does not include a comparable standard concerning feasibility. However, two sections do mention the importance of determining feasible audit objectives (objectives, scope and methodology; 8.09 and understanding the program; 7.10).

The PES practical procedures standard was not identified by the Joint Committee as particularly relevant for reporting an evaluation and is not incorporated into the metaevaluation criteria.

F2 Political Viability

The inclusion of this standard in the PES, and its absence from the GAGAS, illustrates the different contexts of these two professional standards. The GAGAS is designed to direct the actions of government employees who “need to be sufficiently removed from political pressures to ensure that they can conduct their audits objectively and report their findings, opinions, and conclusions objectively without fear of political repercussions” (General Accounting Office, 2003, p. 49). The PES are intended for a much broader audience—an audience who both uses and conducts evaluations. The Joint

Committee writes that the Program Evaluation Standards “have been developed for use by teachers, administrators, school board members, trainers, evaluators, curriculum specialists, legislators, personnel administrators, counselors, community leaders, business and educational associations, parents, and others” (Joint Committee, 1994, p. 1).

The language in this PES standard calling for evaluators to anticipate “the different positions of various interest groups, so that their cooperation may be obtained” shows the value the Joint Committee places on gaining the support of various evaluation stakeholders (p. 71). The GAGAS, in contrast, does not address an auditor’s need to gain support or cooperation. This may display an assumption that a government auditor has an inherent authority that necessarily leads to cooperation.

The PES political viability standard, which was not identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria.

F3 Cost Effectiveness

An evaluation “should be efficient and produce information of sufficient value, so that the resources expended can be justified” (Joint Committee, 1994, p. 77). The PES simply states that an evaluation is cost effective if its benefits are greater than or equal to its costs. Evaluators are to design an evaluation so as to maximize the benefits and minimize the costs. It is not sufficient to merely produce an evaluation design in which benefits outweigh costs, when there are multiple possible designs the one with the best cost-benefit combination should be undertaken.

The GAGAS is silent on the cost-effectiveness of performance audits. This is surprising because much of the purpose for performance auditing is increasing accountability, efficiency, and effectiveness in government operations. If an audit results in decreased costs, or increased efficiencies for an agency or program, any savings need to be measured against the cost incurred to produce the savings (i.e., the cost of the performance audit). However, the topic of cost-effectiveness in the GAGAS is only addressed in terms of the cost-effectiveness of programs and organizations audited.

The PES cost effectiveness standard, which was not identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria because of a lack of information in the reports about the costs of conducting the reports.

P1 Service Orientation

This standard provides that an evaluation should be “designed to assist organizations to address and effectively serve the needs of the full range of targeted populations” (Joint Committee, 1994, p. 83). The GAGAS addresses this service orientation at the level of the auditor instead of the level of the audit itself. Under the section covering auditors’ responsibilities the GAGAS provides that auditors “in discharging their professional responsibilities . . . need to observe the principles of serving the public interest” and “act in a way that will serve the public interest, honor the public trust, and uphold their professionalism” (General Accounting Office, 2003, p. 13-

14) The GAGAS embodies the belief that accountability “is fundamental to serving the public interest” (p. 14).

The PES service orientation standard, which was identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria because of a lack of information in the reports about the service orientation of the reports.

P2 Formal Agreements

The PES requires evaluators and their clients to agree, in writing, to the duties of each party and the process and timeline for the evaluation. This standard, like many PES standards, assumes an independent, outside evaluator working for a client. This model is not assumed under the GAGAS.

The GAGAS standards addressing objectives, scope, and methodology come the closest to paralleling the PES formal agreements standard. However, the GAGAS standards are intended to help “report users to understand the purpose of the audit and the nature of the audit work performed, to provide perspective as to what is reported, and to understand any significant limitations in audit objectives, scope, or methodology” (General Accounting Office, 2003, p. 160). Since government auditor working under the GAGAS are typically working under direct employers/supervisors the need for formal agreements is not as great as in the case of independent consultants working for an organization.

The PES formal agreements standard, which was not identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria because the evaluators in the LPE context are employees of the evaluation funders (the legislators). Information concerning “[o]bligations of the formal parties to an evaluation” is covered in the scope and objectives criteria under preliminaries (Joint Committee, 1994, p. 87).

P3 Rights of Human Subjects

The obligation of evaluators to respect and protect the rights of human subjects in the design and conduct of an evaluation is strongly stated in the PES. The GAGAS does not address human subjects’ rights nearly as strongly. The GAGAS does however outline auditors’ responsibilities to “serve the public interest”, respect an audited entity’s “confidentiality laws, rules, or policies”, and “observe both the form and the spirit of technical and ethical standards” (General Accounting Office, 2003, p. 13, 14, 15).

These GAGAS standards are aimed to ensure the independence and objectivity of the auditors and, consequently, the credibility of the audit findings. This means that the GAGAS is virtually silent on the issue of protecting the human subjects who may be involved with auditors in the collection of data.

The PES rights of human subjects standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) is incorporated into the metaevaluation criteria under the same name in the preliminaries section.

P4 Human Interactions

Human interactions, as identified by the PES, pertain to “evaluators’ interpersonal transactions that affect the feelings and self-respect of those who participate in an evaluation or are affected by an evaluation” (Joint Committee, 1994, p. 99). This is a broader group than human subjects and a broader responsibility than respecting rights. This standard requires evaluators to respect feelings. There is no corresponding standard in the GAGAS.

The PES human interactions standard, which was not identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria.

P5 Complete and Fair Assessment

The program evaluation standards require an evaluation to be “complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated” (Joint Committee, 1994, p. 105). The GAGAS standards cover similar ground, while using slightly different language. The reporting standards for performance audits, 8.41, contend that completeness “requires that the report contain all evidence needed to satisfy the audit objectives and promote an adequate and correct understanding of the matters reported” (General Accounting Office, 2003, p. 174).

Again the GAGAS standards are focused on the scope and objectives of the performance audit while the PES does not mention the evaluation’s goals but instead

concentrates on the strengths and weaknesses of the program. The PES also state that “[e]ven when the primary purpose of an evaluation is to determine the weaknesses of an object, it is essential to identify strengths as well” (Joint Committee, 1994, p. 105).

The PES complete and fair assessment standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) is incorporated into the metaevaluation criteria under the same name in the preliminaries section. The GAGAS complete, 8.41, standard is also incorporated into the metaevaluation criteria under this section.

P6 Disclosure of Findings

The PES requires “the formal parties to an evaluation” to ensure evaluation findings are made “accessible to the persons affected by the evaluation” (Joint Committee, 1994, p. 109). This disclosure requirement is linked to the evaluation’s credibility. The general GAGAS standards require all auditors, when reporting their results, to disclose “all material or significant facts known to them which, if not disclosed, could mislead knowledgeable users, misrepresent the results, or conceal improper or unlawful practices” (General Accounting Office, 2003, p. 15).

Under the same general standard--Auditor’s Responsibilities—the GAGAS calls on auditors to “maintain and broaden public confidence” in auditors and their work. This is another area where the GAGAS and the PES differ slightly. The PES is much more concentrated on public confidence, or stakeholder confidence, in evaluation reports and

evaluation findings. The GAGAS frequently refers to public confidence in auditors as a profession.

The PES disclosure of findings standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) is not incorporated into the metaevaluation criteria because evaluating the extent and effectiveness of disclosure is very difficult using just the evaluation or audit report.

P7 Conflict of Interest

As similar case where the GAGAS focuses on auditors while the PES focuses on processes and results is in the area of conflict of interest. Both the PES (P7) and the GAGAS (auditors' responsibilities, 1.24) address conflicts of interest. The PES requires conflicts of interest to be "dealt with openly and honestly" (Joint Committee, 1994, p. 115). The PES states that conflict exist when "the personal or financial interests of an evaluator might either influence an evaluation or be affected by the evaluation" (Joint Committee, 1994, p. 115). The PES contains several steps for dealing with actual and potential conflicts including clearly identifying possible conflicts in discussions with clients, developing written guidelines for addressing the conflicts, seeking advice from outsiders without these conflicts, arranging for metaevaluations.

The GAGAS states that auditors "should be free of conflicts of interest in discharging their professional responsibilities" and should be "independent in fact and in appearance" (General Accounting Office, 2003, p. 15). The GAGAS makes no arrangements for dealing with conflicts of interest; it completely forbids conflicts of

interest. Auditors covered by the GAGAS are mostly government employees who perform their duties with a degree of independence not usually apparent for evaluators who work directly for clients. This may explain the differences between the two approaches to conflicts of interest.

The PES conflict of interest standard, which was not identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria because of the difficulty in judging the existence of conflicts from the evaluation report or audit.

P8 Fiscal Responsibility

The PES identifies fiscal responsibility on the part of an evaluator as one of the eight “propriety” standards. This standard states that an “evaluator’s allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate” (Joint Committee, 1994, p. 121).

The GAGAS has no standards addressing the fiscal responsibility of auditors in carrying out their duties. A few standards touch on the issue but none address it as directly as does the PES. The external impairments standard (3.19a, p. 44) cautions auditors about factors external to the audit organization that may hinder the objectivity or independence of an audit. One of these factors mentioned is “external interference or influence” that could “reduce inappropriately the extent of work performed in order to reduce costs or fees” (General Accounting Office, 2003, p. 44). The “quality control and

assurance” standard (3.51, p. 57) governs an audit organization’s internal quality control system. Among the factors to be considered in implementing this system is “appropriate cost-benefit considerations” (General Accounting Office, 2003, p. 57). Finally, the standard addressing audit recommendations (8.29, p. 170) requires that auditors’ recommendations be, among other things, cost effective.

The PES fiscal responsibility standard, which was not identified by the Joint Committee as particularly relevant for reporting an evaluation, is not incorporated into the metaevaluation criteria due to the lack of information concerning audit or evaluation cost in almost all LPE reports.

A1 Program Documentation

This PES standard provides that evaluand be “described and documented clearly and accurately, so that the program is clearly identified” (Joint Committee, 1994, p. 127). The GAGAS reporting standards for performance audits, under the accurate standard (General Accounting Office, 2003, p. 174), covers similar ground by requiring “[c]orrect portrayal” of what is being audited.

Again, the PES is most concerned with the affect accurate description has on the usefulness of the evaluation to others while the GAGAS is concerned with accuracy because inaccuracies can damage the credibility of an audit organization.

The PES program documentation standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS accurate standard are

incorporated into the metaevaluation criteria under “accurate” in the methodology section.

A2 Context Analysis

On the issue of context, the PES and the GAGAS are in close agreement. The PES provides that the “context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified” (Joint Committee, 1994, p. 133). The GAGAS “findings” standard (8.13) requires audit reports to “provide selective background information to provide the context for the overall message and to help the reader understand the findings and significance of the issues discussed” (General Accounting Office, 2003, p. 162).

The PES context analysis standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS findings standard are incorporated into the metaevaluation criteria under context analysis in the results section.

A3 Described Purposes and Procedures

This PES standard concerns both the objectives, or purposes, of an evaluation and the procedures used in conducting the evaluation. The A3 standard requires purposes and procedures to be “monitored and described in enough detail, to that they can be identified and assessed” (Joint Committee, 1994, p. 137). The corresponding GAGAS standard is objectives, scope, and methodology (8.08). This standard is intended to allow “report

users to understand the purpose of the audit and the nature of the audit work performed, to provide perspective as to what is reported, and to understand any significant limitations in audit objectives, scope, or methodology” (General Accounting Office, 2003, p. 160).

Both standards in this case are described as necessary for the end use of an audit or evaluation to understand and assess the report.

The PES described purposes and procedures standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS objectives, scope, and methodology standard are incorporated into the metaevaluation criteria under scope and objectives in the preliminaries section.

A4 Defensible Information Sources

This standard requires the sources of “information used in an evaluation” to be “described in enough detail, so that the adequacy of the information can be assessed” (Joint Committee, 1994, p. 141).

The GAGAS does not contain a comparable information sources standard. This may be because the performance auditing standards place an emphasis on concision. The report quality elements standard (8.38) requires reports to be “timely, complete, accurate, objective, convincing, clear, and as concise as the subject permits” (General Accounting Office, 2003, p. 173). While this standard does require the audit itself to be complete and accurate, in other words adequate, it does not require the auditor to describe to the report reader the source of the information.

The PES defensible information standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) is incorporated into the metaevaluation criteria under accurate in the methodology section.

A5 Valid Information

This standard, along with the next four standards, concerns information gathered and analyzed in an evaluation. Together these five standards cover important issues in the process of gathering and analyzing information for an evaluation. Conducting a metaevaluation using only an evaluation, or audit, report necessarily limits the criteria used. None of these five PES standards were identified by the Joint Committee as particularly relevant for reporting an evaluation and none are incorporated into the metaevaluation criteria.

The valid information standard requires information gathering procedures to be developed and implemented to assure that the evaluation's interpretation of the information is "valid for the intended use" (Joint Committee, 1994, p. 145). The PES includes five elements of the "validation process" (p. 145).

The GAGAS addresses the validity of data in various standards. The standard for identifying sources of audit evidence (7.31) requires auditors to "consider the validity and reliability of the data, including data collected by the audited entity, data generated by the auditors, or data provided by third parties, as well as the sufficiency and relevance of the evidence" (General Accounting Office, 2003, p. 143). Another standard, tests of evidence (7.52a), sets the requirement for the sufficiency of data validity in an audit. An auditor

should “ensure that enough evidence exists to persuade a knowledgeable person of the validity of the findings” (General Accounting Office, 2003, p. 150). Finally, the convincing standard (8.49) sets the bar for the level of detail about data validity presented in an audit report. The report should present information that is “sufficient to convince the report users to recognize the validity of the findings” (General Accounting Office, 2003, p. 176).

A6 Reliable Information

The reliable information standard requires information gathering procedures to be chosen, developed and implemented to assure the information is “sufficiently reliable for the intended use” (Joint Committee, 1994, p. 153). In assessing the degree of reliability, evaluators should account for sources of random error and the evaluation’s focus, or target group.

The GAGAS addresses the reliability of data in many of the same standards it addresses the validity of data. The standard for identifying sources of audit evidence (7.31) requires auditors to “consider the validity and reliability of the data, including data collected by the audited entity, data generated by the auditors, or data provided by third parties, as well as the sufficiency and relevance of the evidence” (General Accounting Office, 2003, p. 143).

A7 Systematic Information

This standard requires evaluators to systematically review information collected during the evaluation to ensure that errors are found and corrected. “Prudent evaluators will institute a systematic program of training, controls, and accuracy checks so as to eliminate as many errors as possible and/or to assess the probable effect of the errors that are not detected” (Joint Committee, 1994, p. 159).

The GAGAS uses the term “quality control process” to cover the same ground as the PES’s systematic information control. The quality control and assurance standard (3.49) requires audit organizations to have an “appropriate internal quality control system in place” and to “undergo an external peer review” (General Accounting Office, 2003, p. 56). One quality control process specifically mentioned is referencing. Referencing is “a process in which an experienced auditor who is independent of the audit verifies that statements of facts, figures, and dates are correctly reported, that the findings are adequately supported by the audit documentation, and that the conclusions and recommendations flow logically from the support” (General Accounting Office, 2003, p. 175).

A8 Analysis of Quantitative Information

Quantitative information should be “appropriately and systematically analyzed so that evaluation questions are effectively answered” (Joint Committee, 1994, p. 165).

The GAGAS does not make a clear distinction between quantitative and qualitative information. This may be because the performance auditing field, and the auditing field in general, do not have the same amount of discussion surrounding the issue of quantitative versus qualitative methods as the evaluation field does (see Shadish, 1995; Reichardt & Rallis, 1994; and Reichardt & Cook, 1979).

A9 Analysis of Qualitative Information

This standard sets the same standard for qualitative information that the previous standard set for quantitative information, that is qualitative information should be “appropriately and systematically analyzed so that evaluation questions are effectively answered” (Joint Committee, 1994, p. 171). As stated above, the GAGAS does not make a clear distinction between qualitative and quantitative information.

A10 Justified Conclusions

Standard A10 provides that “conclusions reached in an evaluation should be explicitly justified, so that the stakeholders can assess them” (Joint Committee, 1994, p. 177). The PES cites two reasons for this standard: 1) conclusions that are not justified may be disregarded by the stakeholders, and 2) conclusions that are not justified adequately may be incorrect and lead the evaluation’s “audience to inappropriate actions” (p. 177). An evaluation report must include “full information about how the evaluation

was conducted” and “[w]here possible . . . [a] discussion of plausible alternative explanations of the findings and why these explanations were rejected” (p. 177).

Two GAGAS standards address this issue: evidence (7.48) and convincing (8.49). The convincing standard provides that the “conclusions and recommendations follow logically from the facts presented” (General Accounting Office, 2003, p. 176). The link between a report’s justification and the actions of stakeholders is addressed by GAGAS in a similar way as it is addressed by the PES. Reports that justify their conclusions “help focus the attention of responsible officials on the matters that warrant attention and can provide an incentive for taking corrective action” (p. 176). The evidence standard requires “sufficient, competent, and relevant evidence” to provide a “reasonable basis for the auditor’s findings and conclusions” (p. 149).

The PES justified conclusions standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS evidence and convincing standards are incorporated into the metaevaluation criteria under defensible in the methodology section.

A11 Impartial Reporting

This standard requires an evaluator’s reporting procedures to “guard against distortion caused by personal feelings and biases of any party to the evaluation” (Joint Committee, 1994, p. 181). The evaluator is cautioned to guard against his/her own biases and also against the biases of those providing information for the evaluation.

The GAGAS objective standard (8.46) requires the presentation of evidence in an unbiased manner. The personal impairment standard (3.07) identifies several personal impairments that may affect an auditor's independence. Among these impairments are "preconceived ideas toward individuals, groups, organizations, or objectives of a particular program that could bias the audit" (General Accounting Office, 2003, p. 31).

The PES impartial reporting standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS objective and personal impairment standards are not incorporated into the metaevaluation criteria due to the difficulty of judging these standards based solely on the evaluation or audit report.

A12 Metaevaluation

Standard A12 provides that "the evaluation itself should be formatively and summatively evaluated against these and other pertinent standards" (Joint Committee, 1994, p. 185). The evaluation should be evaluated both by "those who design and carry it through" and by "external metaevaluators" (p. 185). Specific metaevaluations can help evaluators "avoid critical mistakes" and the "[r]egular employment of metaevaluations should enhance the credibility of particular program evaluations and the overall evaluation profession" (p. 185).

The GAGAS does not use the term metaevaluation but it does have standards that address a similar process of quality review of performance audit reports. Standard 1.27, auditor organizations' responsibilities, provides that each audit organization has the responsibility to ensure that "an independent peer review is periodically performed

resulting in an opinion issued as to whether an audit organization's system of quality control is designed and being complied with to provide reasonable assurance of conforming with professional standards" (General Accounting Office, 2003, p. 16). The quality control and assurance standard (3.49) also requires periodic external peer review.

The PES metaevaluation standard (identified by the Joint Committee as particularly relevant for reporting an evaluation) and the GAGAS audit organizations' responsibilities and quality control and assurance standards are incorporated into the metaevaluation criteria under quality control in the preliminaries section.

Conclusions: PES and GAGAS Standards

Most of the differences between the PES and the GAGAS can be understood by the differences in audience, creation, and contexts of use.

The main audience for the GAGAS is the professional auditors working in a government context. The audience for the PES is much broader. Though focused on an educational context, the intended audience reaches beyond just evaluators to those who use, read, and fund evaluation reports. This difference leads the PES to be much more inclusive and receptive to the views of stakeholders beyond the evaluators. The views of program/agency staff, consumers and other stakeholders are consistently recognized in the PES. The GAGAS references broader public interests in introductory statements about an auditor's responsibility but these references are not followed in the text of the various, specific standards.

The creation of the PES was a highly collaborative process that arose from meetings of several educational and professional organizations. While the original GAGAS, and especially its subsequent revisions, were created with a high degree of input it was and is at its core a publication of the U.S. government's General Accounting Office (now the Government Accountability Office). This leads to the GAGAS being much more focused on the needs and issues of government auditors. The PES, in contrast was formed to aid not just evaluators but those who are affected by, fund and use evaluations.

The PES was created within a specific policy area—training and education—while the GAGAS was created in a broad organizational area—government organizations, programs, activities, and functions. However, it is the PES that has broken beyond its specific policy area to reach into the evaluation of many different programs—whether government, nonprofit, or business. The GAGAS remains focused on performance auditing in the government sphere. This shows in the areas of the PES standards such as stakeholder identification, service orientation, rights of human subjects, and human interactions where the GAGAS is silent.

Summary and Conclusions

On the topic of state legislative program evaluation, two discoveries stand out from a review of the literature. First, the terminology varies greatly though over the past several years the terms *performance audit* and *program evaluation* have begun to predominate. As described above, this study uses the term legislative program evaluation

(LPE) to refer to all of these activities. Second, the main methodologies employed when examining LPE is case studies (when looking at LPE reports) and surveys (when looking at LPE unit staffing and structure)

Only two published papers look directly at a sample of state evaluation-audit reports. Both of these papers were published during the 1980s (Brown & Craft, 1980; Brown, 1988). Neither paper uses metaevaluation criteria. Though, Brown and Craft (1980) do identify several descriptive criteria. This lack of academic attention to LPE reports is one inspiration for the structure of the current study.

The metaevaluation literature is more extensive, though not voluminous. Ten published papers apply metaevaluation to multiple evaluation reports (see Table 1, p. 37). Unlike the current study, only one of these papers selected the reports under study randomly (Hoogerwerf, 1992). However, nine of the ten papers used only the report itself to complete the metaevaluation. This study seeks to build on the metaevaluation work of these previous papers.

CHAPTER III

METHODOLOGY

A methodology directly examining the products of legislative program evaluation—the reports themselves—seems to be crucial to get a good idea of what is actually happening in the field. However, very little of this approach has been undertaken. Friedberg and Lutrin (2004) conducted searches of academic journals using several different search engines and found only two recent academic papers concerning state-level auditing. Funkhouser (2005) also cites the scant literature on performance auditing. He sets an agenda for scholarly analysis of performance auditing including: content analysis of actual audit reports, examination of impact of audit organizations, and an examination of the differences in process, products and results in audit organizations that arise from different structures. This study concentrates on two of these areas—an analysis of actual performance audit and evaluation reports and an examination of the different products (reports) that arise from different organizational structures. By focusing on two of Funkhouser’s agenda items the study seeks to contribute to the substantive knowledge of how evaluation is conducted by an important U.S. governmental sector, state legislatures, on a wide range of issues (education, transportation, safety, etc.). The study also aims to contribute to the methodology of evaluation by showing how metaevaluation can be used to assess a large number of evaluation reports in a setting, field or topic.

Legislative Program Evaluation Units

State LPE offices were identified starting with the membership of the National Legislative Program Evaluation Society (NLPES). NLPES is a staff section of the National Conference of State Legislatures. The NLPES identifies 42 LPE units in 41 different states (VanLandingham, Kinton, Boerner, & Zelio, 2000). Texas has two LPE units—the State Auditor’s Office and the Sunset Advisory Committee. The present study included only state LPE units that are directly under the control of the legislature. This includes units where the director is appointed or elected by the legislature, or a committee of the legislature, and can be removed only by the legislature, not by the state’s governor or another executive branch official. This criterion removed two LPE units (from California and Oregon). Further states were eliminated because a review of their audits and reports from 2001 through 2005 showed they were either financial audits or otherwise not related to performance auditing or evaluation (Alabama, Iowa, North Carolina, and North Dakota). This left 36 LPE units from 35 states from the members of NLPES. An investigation of the remaining 15 state legislatures identified five more units (in California, Massachusetts, New York, Oregon, and Rhode Island) that perform evaluations, performance audits, or related reviews¹. A total of 41 LPE units from 40 states were included in this study² (see Appendix B).

¹ NLPES lists the Oregon Secretary of State’s Audits Division and the California State Auditor as members. This study includes the Oregon Legislative Fiscal Office as Oregon’s LPE unit and the California Legislative Analyst’s Office as California’s LPE unit.

² In 2007 NLPES listed members from 58 units in 49 states. Among the 17 units and 9 states not represented in this study are 5 units that conduct all fiscal audits, 5 executive branch units, 4 units that produced no LPE reports from 2001 through 2005, 2 units that were established after 2005, and 1 division within a LPE unit included in this study.

All reports (audits, evaluations, sunset reports, and others) from these 41 units released from January 1, 2001 through December 31, 2005 were identified through reviewing the various unit websites and requesting reports directly from unit staff. Many of the LPE units also perform financial audits. These financial audits were not included in the population of reports from which a sample was drawn. A few states conduct combined financial/performance audits³. These reports were included in the population. A total of 1,911 reports were identified during the five year period under review.⁴

Each report was first classified by how it was identified by each unit (i.e. was it called a performance audit, an evaluation, a review). Then reports were re-classified to include all types of evaluations (performance evaluations, program evaluations) as simply evaluations and all types of audits (performance audits, management audits, program audits, effectiveness audits) as performance audits. Reports referred to as sunset reports were classified as such. Sunset reports came solely from the states of Alaska (36 reports), Arizona (1 report), Texas (44 reports), and Washington (1 report). Reports identified by other names were classified as other. These report names included assessment, examination, investigation, report, review, study, and survey.

Next a random sample of 100 reports was selected by assigning each report a random number between zero and one and choosing the 100 lowest numbers for the sample. The following procedure was used to ensure that the sample included at least one report from each state LPE unit. The original random sample of 100 reports was examined to count the number of LPE units represented. LPE units from nine states

³ From 2001 through 2005 Illinois conducted "Financial and Management Audits" and "Financial, Compliance and Program Audits" while Pennsylvania conducted "Financial and Management Reviews.

⁴ Reports were identified, collected, and analyzed during 2006. During 2007 this analysis was refined and completed.

(Connecticut, Idaho, Kentucky, Maryland, Missouri, New Hampshire, Nevada, Oregon, and Pennsylvania) were not represented in these 100 reports. The report from each of these nine states with the lowest randomly generated number among that state's reports was added to the sample, replacing the nine reports with the highest of the original 100 random numbers. These replaced reports came from Georgia (2 reports), Montana, New Mexico, Texas (Sunset Advisory Commission), Utah, Virginia, and West Virginia (2 reports).

Table 3 shows the characteristics of the population of LPE reports identified for the five-year period of 2001 through 2005 and the sample selected from this population. Other reports are slightly overrepresented in the sample while performance audits are slightly underrepresented. Similarly, one year (2002) is slightly overrepresented while another year (2001) is slightly underrepresented. As the random sample was stratified by LPE unit, all 40 states and all 41 LPE units are represented in the sample.

Table 3. Reports Sample and Population.

| Report type | Sample (%) | Population (%) |
|-------------------|------------|----------------|
| Other | 23 (23.0) | 344 (18.0) |
| Evaluation | 25 (25.0) | 485 (25.4) |
| Performance audit | 47 (47.0) | 1,000 (52.3) |
| Sunset | 5 (5.0) | 82 (4.3) |
| Report year | | |
| 2005 | 19 (19.0) | 366 (19.2) |
| 2004 | 19 (19.0) | 371 (19.4) |
| 2003 | 18 (18.0) | 360 (18.8) |
| 2002 | 26 (26.0) | 410 (21.5) |
| 2001 | 18 (18.0) | 404 (21.1) |

Finally, a metaevaluation was conducted on all 100 sample reports using criteria (see Appendix A) adapted by combining elements of the U.S. Government Accountability Office's *Generally Accepted Government Auditing Standards* for

performance audits (GAGAS), the Joint Committee on Educational Evaluation's *Program Evaluation Standards* (PES), and Scriven's *Key Evaluation Checklist* (KEC).

Non-legislative Program Evaluation Units

To provide a comparison of different organizational structures non-legislative program evaluation units (i.e. units that were not under the control of the legislative branch) were identified that exist in the ten states that did not have a LPE unit issuing reports during 2001 through 2005 (Alabama, Delaware, Iowa, Maine, New Jersey, North Carolina, North Dakota, Oklahoma, South Dakota, and Vermont). Four of these states (see Table 4) had an executive branch agency issuing performance audit or evaluation reports during 2001 through 2005.

Table 4. State Non-legislative Program Evaluation Units.

| State | Non-LPE Unit | Description |
|----------------|-----------------------------|--|
| North Carolina | Office of the State Auditor | Unit performs mostly financial audits and some performance audits. |
| North Dakota | State Auditor | Unit performs mostly financial audits and some performance audits. |
| Oklahoma | State Auditor and Inspector | Unit performs mostly financial audits and some performance audits. |
| Vermont | State Auditor | Unit performs mostly financial audits and some special audits including special reviews, program assessments and program audits. |

These four units issued a total of 53 reports from 2001 through 2005. These 53 reports were classified as either performance audits or other reports. A sample of 10 non-LPE reports was selected using the same procedure described to select the LPE report sample. Table 5 shows information concerning the non-LPE report population and

sample. The sample of non-LPE reports was metaevaluated using the same criteria applied to the LPE report sample.

Table 5. State Non-legislative Program Evaluation Reports.

| Report type | Sample (%) | Population (%) |
|-------------------|------------|----------------|
| Other | 2 (20.0) | 18 (34.0) |
| Evaluation | 0 (0.0) | 0 (0.) |
| Performance audit | 8 (80.0) | 35 (66.0) |
| Sunset | 0 (0.0) | 0 (0.0) |
| Report year | | |
| 2005 | 2 (20.0) | 13 (24.5) |
| 2004 | 3 (30.0) | 13 (24.5) |
| 2003 | 1 (10.0) | 13 (24.5) |
| 2002 | 2 (20.0) | 10 (18.9) |
| 2001 | 2 (20.0) | 4 (7.5) |
| Report state | | |
| North Carolina | 4 (40.0) | 16 (30.2) |
| North Dakota | 2 (20.0) | 7 (13.2) |
| Oklahoma | 1 (10.0) | 8 (15.1) |
| Vermont | 3 (30.0) | 22 (41.5) |

Report Recommendations

To investigate the mechanisms by which the various units seek to affect policy change report recommendations were analyzed. A specific question in this regard is whether the Louisiana Legislative Auditor used its unique statutory authority to evaluate ““basic assumptions underlying any and all state agencies and the programs and services provided by the state”. (Louisiana Performance Audit Program Act, 1995) A review of each state’s statute that creates and governs its LPE unit revealed Louisiana to be the only state specifically granting its LPE unit this power to question the legislature’s wisdom in creating an agency or program. Recommendations issued by LPE unit reports, non-LPE unit reports, and Louisiana reports were analyzed.

In addition to comparing different report categories—other, evaluation, performance audit, and sunset—LPE unit reports were compared to non-LPE reports. Reports from the Louisiana LPE unit were also examined separately due to their unique statutory authority.

All reports were examined for the presence of recommendations. Most reports include recommendations under a separate heading and number the recommendations. In this situation separate recommendations were counted according to the numbering system in the report. In many instances a single numbered recommendations included several sub-points. However, this set of recommendations was counted as a single recommendation in keeping with the report's numbering system. Some reports include unnumbered recommendations. In these instances separate recommendations were counted based on how they were presented in the text.

Each recommendation was coded as either a recommendation for an executive agency (an agency recommendation) or a recommendation for the legislature (a legislative recommendation). Each recommendation was further coded as meeting one or more of the following criteria: (i) requesting further study of an issue (either by the unit writing the report, by the legislature, or by some other organization), (ii) requesting a clarification from the agency or the legislature in the rules or laws concerning the issues involved, (iii) making a substantive change in the programs, policies, or processes of the agency involved, (iv) requesting the elimination of a program, agency, duty, or power, or (v) requesting the consolidation of certain organizations or duties into one or more entities.

Louisiana Subsample

To examine the extent to which the Louisiana Legislative Auditor exercised its statutory authority to evaluate the “basic assumptions underlying any and all state agencies and the programs and services provided by the state” (Louisiana Performance Audit Program Act, 1995) a random sample of ten of the 42 performance audits released from 2001 through 2005 were metaevaluated using the same metaevaluation criteria applied to other states’ LPE reports and non-LPE reports.

Metaevaluation Criteria

The GAGAS standards are the most commonly applied standards for the reports produced by the state LPE units (see below). The PES are far less commonly cited but they are commonly used for metaevaluation in the evaluation literature. (Chatterji, 2005; Sanders, 1999; Scott-Little, Hamann, & Jurs, 2002)

For both ease of presentation and feasibility of applying metaevaluation criteria to a large sample of reports three main categories—preliminaries borrowing a term from Scriven’s [2006] Key Evaluation Checklist), methodology, and results—with 16 total criteria were chosen. The preliminaries section includes criteria for statement of scope and objectives, identification of values, quality control, statement of standards compliance, and stakeholder identification. The methodology section includes criteria for completeness, accuracy, defensibility, human subjects’ rights, and comparisons. The results section includes criteria for context analysis, clarity, concision, findings and

conclusions, recommendations, and timely. Appendix A shows each of the 16 metaevaluation criteria and descriptions of classifications for *excellent*, *good*, *fair* and *poor*.

Some metaevaluation systems (Scott-Little, Hamann, & Jurs, 2002; Stufflebeam, 2000b; Stufflebeam, 2001a; Stufflebeam, 2001b) use more grades than the simple dichotomy of *meets* and *fails to meet*. The ratings system adopted for this study seeks to add detail to the evaluation of each criterion without creating overly specific categories that are very difficult to differentiate in practice. In general, excellent and good ratings can be seen as meeting a criterion. Fair ratings are partially meeting a criterion. Poor ratings are failing to meet a criterion. Not addressed means a grade for that criterion cannot be determined from the information presented in the report. However, on some criteria, a report's failure to address the criterion results in a grade of poor because that criterion should be addressed in the text of a report (e.g., scope and objectives).

Each report was reviewed by the author, who applied the metaevaluation criteria and assigned the grades on each of the 16 criteria. Each report was read through first without assigning grades. During a second reading the author assigned grades for each criterion and entered justifications for each grade into a spreadsheet. Each report was read a final time to confirm or change the grades assigned for each criterion. The first and third readings of each report took approximately 1 hour each. The second readings, during which grades were assigned, took approximately 3 hours each (including time spent recording grades and notes on each grade). Reports varied greatly in their lengths, so these time estimates are averages used to show the approximate amount of time spent reviewing each report. Report page totals can be deceiving due to the inclusion or

exclusion of various appendices or supporting data. However, the average page length of the sample LPE reports was 50.97.

Table 6. Metaevaluation Criteria and Sources.

| Metaevaluation criterion | GAGAS | PES | KEC |
|----------------------------|--|--|--|
| Scope and objectives | Objectives, scope, and methodology (8.08) | Described purposes & procedures (A3) | Preface (Preliminaries II) |
| Values identified | | Values identification (U4) | Values (Foundations 5) |
| | Audit organizations' responsibilities (1.27) | | |
| Quality control | Quality control and assurance (3.49) | Metaevaluation (A12) | Methodology (Preliminaries III) |
| Standards compliance | Statement on GAGAS compliance (8.30) | | Metaevaluation (Conclusions and implications 15) |
| Stakeholder identification | | Stakeholder identification (U1) | Preface (Preliminaries II) |
| Complete | Complete (8.41) | Information scope & selection (U3) | |
| Accurate | Accurate (8.43) | Accuracy (A1, A4) | |
| Defensible | Evidence (7.48) Convincing (8.49) | Justified conclusions (A10) | |
| Human subjects' rights | | Rights of human subjects (P3) | Metaevaluation (Conclusions and implications 15) |
| Comparisons | Evidence (7.50) | | Comparisons (Subevaluations 9) |
| Context analysis | Findings (8.13) | Context analysis (A2) | Preface (Preliminaries II) |
| Clear | Clear (8.50) | Report clarity (U5) | |
| Concise | Concise (8.53) | | |
| Findings and conclusions | Findings (8.13) Conclusions (8.27) Convincing (8.49) | | Synthesis (Conclusions and implications 11) |
| Recommendations | Recommendations (8.28) | | Recommendations and explanations (Conclusions and implications 12) |
| Timely | Timely (8.39) | Report timeliness and dissemination (U6) | |

Table 6 shows the metaevaluation criteria and the corresponding sources for each. All the PES standards cited above were identified by the Joint Committee as being particularly important when “reporting the evaluation” (Joint Committee on Standards for Educational Evaluation [Joint Committee], 1994, p. ix). The absence of a citation to one of the three sources does not mean that source does not require the criteria, simply that it is not a specifically stated criterion (for example, there is no specific citation in the Key Evaluation Checklist to accuracy).

Developing criteria to evaluate an evaluation report necessarily misses some important aspects of the overall evaluation. Many aspects of the underlying evaluation are either not fully addressed in the report or cannot be accurately judged by reviewing just the report. For instance, a report may indicate steps taken to uphold the rights of human subjects during the course of the evaluation. The contents of the evaluation report are not sufficient to determine if these steps were actually taken during the course of conducting the evaluation.

For these reasons, Table 7 shows the 16 metaevaluation criteria and the degree to which each criterion is able to be accurately assessed based on the evaluation report alone.

Table 7. Confidence in Metaevaluation Criteria.

| Metaevaluation criterion | Degree of confidence | Rationale |
|----------------------------|----------------------|--|
| Scope and objectives | High | Scope and objectives are easily reported in the evaluation report |
| Values identified | Moderate | Values are written about in many ways. Not always clearly identified. Often implied. |
| Quality control | Low | Quality control steps can be listed but verifying their use is impossible using just the report. |
| Standards compliance | High | A statement of compliance with standards is easily reported in the evaluation report. |
| Stakeholder identification | Moderate | Stakeholders are written about in many ways. They are not always clearly identified as stakeholders. |
| Complete | Moderate | Completeness can be judged against the goals of the report, but it can be difficult to decide if side-effects and other issues are ignored. |
| Accurate | Moderate | Accuracy can be checked by agency comments. However, these comments are not always included and when included are often only good checks against negative conclusions. |
| Defensible | Moderate | Arguments and evidence can be assessed within the report. However, data often cannot be confirmed. |
| Human subjects' rights | Moderate | Procedures for addressing human subjects' rights can be identified in the report, but the evaluators' practices cannot. |
| Comparisons | Moderate | Comparative studies and programs/agencies are easily identified in the report. However, the depth of the comparison made is often not fully addressed. |
| Context analysis | Moderate | Context provided in reports can be judged but additional contextual issues that may have been missed cannot be judged solely using the report. |
| Clear | High | The clarity of the report can be judged. |
| Concise | Moderate | Concision of the report can be judged but given the differences in topic complexity, importance, etc. it is highly subjective. |
| Findings and conclusions | Moderate | Evidence supporting the findings can be judged by the report, but data often cannot be confirmed. |
| Recommendations | Moderate | Recommendations can be judged from their presentation in the report. However, data supporting the recommendations often cannot be confirmed. |
| Timely | Low | Information on timeliness is often not included in the evaluation reports. |

Some of the metaevaluation criteria are useful for evaluating a report in very limited ways. The accuracy criterion can only be used to confidently determine if a report is inaccurate in ways unfavorable to program or agency staff. In many reports, these staff members have the opportunity to comment on the evaluation and to protest errors. It is likely the staff members will not protest errors that present their program or agency in a favorable light.

Other criteria only address limited aspects of a report's merit. The scope and objectives criterion only concerns specificity and provenance of the report's scope and objectives. This fails to address important topics such as the appropriateness of the evaluation's scope (too broad or too narrow), the significance of the evaluand (is it worth the effort used to evaluate it), and the feasibility of the scope of the project given the resources provided (time, money, staffing).

To develop a metaevaluation scheme that can work for LPE, be replicated across different sets of reports, and brief enough to be feasible for large samples of reports 16 criteria were chosen, broken into three categories. These criteria were developed to closely conform to the language from the GAGAS and the PES. Indeed, much of the specific language reflected in each criterion come from one or both of these sources. Following is a detailed examination of each criterion highlighting the attributes of reports graded excellent, good, fair, and poor on each specific criterion.

Preliminaries

Criteria in this category include statement of scope and objectives, identification of values, quality control, statement of standards compliance, and stakeholder identification. These sections of a report introduce the topic under review, identify the focus areas of the report, and give general guidelines of how the report was conducted.

Statement of Scope and Objectives

The general GAGAS standard for report contents (8.07) requires reports to include “the objectives, scope, and methodology . . .” (General Accounting Office, p. 160). Points 8.08 through 8.12 further refine what should be included regarding objectives, scope, and methodology. Objectives should be stated in a “clear, specific, and neutral manner”, “explain why the audit organization undertook the assignment and state what the report is to accomplish and why the subject matter is important” (p. 160). The report should “clearly describe the scope of the work performed and any limitations; any applicable standards that were not followed” and why (p. 161).

Scriven’s Key Evaluation Checklist (KEC) notes that the preface is the place to define “[e]xactly what you’re supposed to be evaluating (what’s the evaluand)?” (Scriven, 2006, p. 2) Scriven also suggests the evaluator specify whether she will be evaluating an entire program, specific dimensions of a program, or just certain components of the program. All these points hew closely to the GAGAS scope requirements while they are somewhat related to the objectives requirements.

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report lists the specific agency/department and/or activities to be evaluated/audited. The report lists the specific objectives of the audit/evaluation and how the objectives were determined (by legislative direction, statute, etc.).

Good: The report lists the specific agency/department and/or activities to be evaluated/audited. The report lists the specific objectives of the audit/evaluation but not how the objectives were determined.

Fair: The report specifies the agency/department/activity evaluated or the specific objectives but not both or lists nonspecific objectives (such as “operating in the public interest”).

Poor: The report fails to specify both the agency/department/activities evaluated/audited and the objectives.

Identification of Values

The GAGAS do not address values at all. In fact a search of the GAGAS finds not use of the word “values.” The only instances of the word “value” pertains to monetary values, such as in defining fraud as obtaining something of value through misrepresentation.

The PES and the KEC each stress the importance of values.

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report clearly identifies the values used to determine the evaluative conclusions and how these values were chosen (the source of these values).

Good: The report clearly identifies the values used to determine the evaluative conclusions but not how these values were chosen (the source of these values).

Fair: The report implies the values used without clearly identifying them.

Poor: The report fails to address the values used.

Quality Control

The GAGAS includes a standard (3.49) for “quality control and assurance.” This standard specifically addresses the organization performing the audit rather than the performance of the audit itself. The standard requires each organization to “have an appropriate internal quality control system in place” and to “undergo external peer review” (General Accounting Office, p. 56).

The KEC does not use the language of quality control or external peer review. However, the methodology checkpoint addresses issues such as ensuring the evaluation team has “adequate domain expertise” through the use of consultants or advisory panels (Scriven, 2006, p. 2).

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report clearly identifies sufficient internal quality control measures and external peer reviews conducted of its organization.

Good: The report clearly identifies sufficient internal quality control measures but fails to identify external peer reviews conducted of its organization.

Fair: The report identifies limited internal quality control measures and no external organizational reviews.

Poor: The report identifies no internal quality control measures and no external organizational reviews.

Statement of Standards Compliance

The GAGAS clearly require that “[a]uditors should report that the audit was made in accordance with GAGAS” (General Accounting Office, p. 170). There is no requirement as to the form of this report, but this standard does specify that any specific standards that were not followed should be reported in the scope section. The KEC indirectly addresses the use of standards by suggesting the use of the PES in conducting a metaevaluation.

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report prominently states what standards (PES, GAGAS), if any, guide its conduct (for example, if there is a methodology section, the standards followed are listed there as well as in the introduction, summary, or letter of transmittal).

Good: The report states what standards, if any, guide its conduct but do not identify these standards in the methodology section.

Fair: The report refers to adherence to general professional standards or “government auditing standards” but does not specify which standards.

Poor: The report does not address standards.

Stakeholder Identification

The first PES utility standard (U1) is “stakeholder identification.” This standard states that “[p]ersons involved in or affected by the evaluation should be identified, so their needs can be addressed” (Joint Committee, p. 25). This standard is particularly important to evaluation reports as the utility standards as a whole are intended “to ensure that an evaluation will serve the information needs of intended users” (p. 23).

The KEC, under the preface section, notes that an evaluator should “identify and define” stakeholders. Stakeholders are defined as “those with a vested interest in the outcome of the evaluation” (Scriven, 2006, p. 2).

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report clearly identifies stakeholders for the program/agency/policy under review, including all of the following: consumers of the evaluand, staff working on the evaluand, decision-makers about the evaluand and funders of the evaluand.

Good: The report clearly identifies stakeholders for the program/agency/policy under review, including more than one but not all of the following: consumers of the evaluand, staff working on the evaluand, decision-makers about the evaluand and funders of the evaluand.

Fair: The report identifies limited stakeholders or identifies only an overly broad group of stakeholders (e.g. the people to the state).

Poor: The report fails to identify any stakeholders.

Methodology

This category includes criteria for completeness, accuracy, defensibility, human subjects' rights, and comparisons. Reports were evaluated on these criteria to judge how well each report was conducted and how that conduct is reflected in the written report.

Complete

This is the only metaevaluation standard that is taken solely from one of the three sources, the GAGAS. The applicable GAGAS standard (8.41) requires “that the report contain all evidence needed to satisfy the audit objectives and promote an adequate and correct understanding of the matters reported” (General Accounting Office, p. 174). Report users should be given an “adequate and correct understanding” by “providing perspective on the extent and significance of reported findings” (p. 174). The subsequent standard (8.42) provides that “[s]ufficient detailed supporting data should be included to make convincing presentations” (p. 174).

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report covers all of the areas it claims to cover in the scope and objectives sections.

Good: The report covers all but a single small area mentioned in the scope and objectives section.

Fair: The report covers most areas it claims to cover in the scope and objectives section.

Poor: The report fails to cover most of the areas it claims to cover in the scope and objectives section.

Accurate

The GAGAS provides specific standards (8.43, 8.44, and 8.45) concerning accuracy. Audit reports “should include only information, findings, and conclusions that are supported by sufficient, competent, and relevant evidence in the audit documentation” (p. 174). These GAGAS standards also stress the damage that can be done not only the specific audit containing inaccuracies but also to the organization that produces the audit.

One of the four categories of PES standards is accuracy. Within the twelve substandards for accuracy are two which are particularly relevant to the metaevaluation criteria used here. Program documentation (A1) provides that the evaluand “should be described and documented clearly and accurately, so that the program is clearly identified” (Joint Committee, p. 127). Defensible information sources (A4) requires information sources to be “described in enough detail, so that the adequacy of the information can be assessed” (p. 141).

GAGAS standard 8.31, “Reporting Views of Responsible Officials” (General Accounting Office, p. 171), requires audits to report the views of those who administer the program under review. Typically comments are received prior to the final audit being issued. These comments are included in an appendix to the audit report. These comments are used to operationalize the accurate criterion for the metaevaluations because checking

the accuracy of information in each audit was not possible working just from the reports themselves.

These standards were translated to the metaevaluation criteria as follows.

Excellent: Agency comments are included and the agency did not dispute the accuracy of the information reported.

Good: The agency disputed the accuracy of some information but the report convincingly rebuts this.

Fair: The agency convincingly disputes the accuracy of some minor information in the report.

Poor: Agency comments convincingly dispute report's accuracy.

Defensible

The GAGAS evidence standard (7.48) reads: “[s]ufficient, competent, and relevant evidence is to be obtained to provide a reasonable basis for the auditors’ findings and conclusions” (p. 149). Concerning the sufficiency of the evidence, standard 7.52(a) provides that the level of evidence is sufficient if “enough evidence exists to persuade a knowledgeable person of the validity of the findings” (p. 150). The PES standard “Justified conclusions” (A10) provides closely related language, stating that an evaluation’s conclusions should be “explicitly justified, so that stakeholders can assess them” (Joint Committee, p. 177).

These standards were translated to the metaevaluation criteria as follows.

Excellent: There was “sufficient, competent and relevant evidence . . . obtained to provide a reasonable basis for the auditors’ findings and conclusions.”

Good: There was “sufficient, competent and relevant evidence . . . obtained to provide a reasonable basis for the auditors’ findings” but the overall conclusions go beyond the individual findings.

Fair: There was competent and relevant information presented but the conclusions go beyond what that evidence supports.

Poor: There was not sufficient, competent and relevant information presented to provide a reasonable basis for the findings.

Human Subjects’ Rights

One of the four main areas of the PES is propriety. The propriety standards are designed to “ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results” (p. 81). One of the substandards for propriety is “Rights of Human Subjects” (P3). P3 provides that “[e]valuations should be designed and conducted to respect and protect the rights and welfare of human subjects” (p. 93).

The KEC briefly touches on this subject by including ethicality among the five “primary criteria of merit for evaluations” (Scriven, 2006, p. 15).

A caution with the operationalization of this metaevaluation criterion is that is merely addressing the presence (in the report) of an obvious violation of human subjects’

rights. Using the reports themselves there is necessarily no good way to determine if the report's authors respected the rights of human subjects.

These standards were translated to the metaevaluation criteria as follows.

Excellent: There is no evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.) and the report identifies procedures designed to ensure the rights of human subjects.

Good: There is no evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.).

Fair: There is some evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.).

Poor: There is strong evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.).

Comparisons

The KEC is the main impetus behind the inclusion of the comparisons criterion. Comparisons are included as a subevaluation, one of the “five core dimensions” of an evaluation along with process, outcomes, costs, and generalizability (p. 10). The key to finding appropriate comparisons is to look for “critical competitors” which are “entities that are alternative ways of getting the same or similar benefits from about the same resources” (p. 10). The KEC also presents a point that is especially relevant to LPE units. In looking for critical competitors “there are also sometimes strong reasons to compare the evaluand with a demonstrably possible alternative—a “virtual criticom”—that could

be assembled from existing or easily constructed components” (p. 10). This does raise some cautions as this moves the evaluator into “the role of a program designer rather than evaluator, which creates a risk of conflict of interest” as the evaluator seeks to compare the original evaluand with a proposed program of her own design. (p. 10)

The GAGAS standards, under evidence (7.50), includes comparisons as a form of analytical evidence but this standard does not require, or even encourage, the use of comparisons.

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report competently addresses evaluations/audits of similar programs/agencies.

Good: The report refers to evaluations/audits of similar programs/agencies but doesn't address them fully.

Fair: The report addresses similar programs/agencies but doesn't address evaluations/audits of these programs.

Poor: The report fails to refer to evaluations/audits of similar programs/agencies.

Results

The results section includes criteria for context analysis, clarity, concision, findings and conclusions, recommendations, and timely. These criteria focus on the merit of the report as presented. Context analysis is included in this category due to its importance in determining how well the policy or program under review may be transferred to different geographic, social, or political contexts.

Context Analysis

The GAGAS only briefly touch on an evaluand's context. The findings standard (8.13) provides that an "audit report should provide selective background information to provide the context for the overall message and to help the reader understand the findings and significance of the issues discussed" (General Accounting Office, p. 163). Both the PES and KEC feature context much more prominently.

Context analysis is a separate substandard (A2) under the PES accuracy standard. This standard provides that the program's context be "examined in enough detail, so that its likely influences on the program can be identified" (Joint Committee, p. 133).

The KEC addresses context in several places but most specifically under the preface section of the preliminaries checkpoint. Questions are posed for evaluators to answer in developing the evaluation. Questions such as: "How much of the context is to be included? How many of the details are important (enough to replicate the program elsewhere, enough to recognize it anywhere, or just enough for prospective readers to know what you're referring to)?" (Scriven, 2006, p. 2).

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report presents the context in which the program/policy/agency exists "in enough detail, so that its likely influences on the program can be identified."

Good: The report presents the context in which the program/policy/agency exists but does not address the context's influences.

Fair: The context is presented in insufficient detail.

Poor: The report doesn't present, or only marginally presents, the context.

Clear

The GAGAS standards include a section regarding report clarity (8.50). This standard provides that “the report be easy to read and understand,” be “prepared in language as clear and simple as the subject permits,” and that “technical terms, abbreviations, and acronyms . . . should be clearly defined” (General Accounting Office, p. 176). A subsequent standard provides that “[l]ogical organization of material, and accuracy and precision in stating facts and in drawing conclusions, are essential to clarity and understanding” (p. 177).

The PES substandard U5, Report Clarity, provides that reports should “clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood” (Joint Committee, p. 49). Like the GAGAS standard for clarity, this PES standard is linked to the utility of the report.

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report is (1) easy to read and understand, (2) technical terms, abbreviations and acronyms are explained, and (3) the material is logically organized.

Good: The report is easy to read and understand and logically organized but technical terms, abbreviations and acronyms are not explained.

Fair: The report is easy to read and understand but is not logically organized and technical terms, abbreviations and acronyms are not explained.

Poor: The report is not easy to read and understand.

Concise

GAGAS standard 8.53 provides that “[b]eing concise requires that the report be no longer than necessary to convey and support the message.” This standard also cautions that “needless repetition should be avoided.” These standards are tied to a report’s utility in that reports that “are fact based but concise are likely to achieve greater results” (General Accounting Office, p. 177).

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report is only as long as is necessary to convey and support conclusions. Repetition and extraneous detail are avoided.

Good: The report is only as long as is necessary to convey and support conclusions but includes some repetition and extraneous detail.

Fair: The report is longer than necessary to convey and support conclusions.

Poor: The report includes extensive repetition and/or extraneous detail.

Findings and Conclusions

As findings and conclusions are closely related they are combined in one metaevaluation criterion here. The GAGAS provides at least three standards directly related to the findings and conclusions criterion—findings (8.13), conclusions (8.27), and convincing (8.49). Standard 8.13 ties findings directly to the audit’s objectives. Standard 8.27 defines conclusions as “logical inferences about the program based on the auditors’ findings and should represent more than just a summary of the findings.” They should be

“clearly stated, not implied” and they are stronger if they “set up the report’s recommendations and convince the knowledgeable user of the report that action is necessary” (p. 169). The convincing standard (8.49) further provides ties to the audit’s objectives. An audit is convincing when its results are “responsive to the audit objectives,” its findings are “presented persuasively,” and its “conclusions and recommendations follow logically from the facts presented” (p. 176). Unlike the KEC, nothing in the GAGAS ties findings or conclusions to consumer/impactee needs.

The KEC addresses findings and conclusions in the synthesis checkpoint under the conclusions and implications heading. Conclusions should focus on “the present and future impact on consumer and community needs, subject to the constraints of ethics and the law, and feasibility, etc.” but there usually “should also be some conclusion(s) that are aimed at the client’s and other stakeholders’ needs for concise evaluative information.” This requires “referencing the results against their goals, wants, or hopes (if feasible) . . . [b]ut the overarching obligation of the evaluator is to reference the results to the needs of the impacted population” (Scriven, 2006, p. 11).

These standards were translated to the metaevaluation criteria as follows.

Excellent: Findings are supported by credible evidence related to the report’s objectives. Conclusions are reported as “logical inferences from the findings” not simply a restatement of findings.

Good: Findings are supported, but some conclusions are simply a restatement of findings.

Fair: Findings are supported, but conclusions are mostly a restatement of findings.

Poor: Findings are insufficiently supported and conclusions are mostly a restatement of findings.

Recommendations

The GAGAS standard for recommendations (8.28) provides that “[i]f warranted, auditors should make recommendations for actions to correct problems identified during the audit and to improve programs and operations.” Recommendations should be made when “the potential for improvement in programs, operations, and performance is substantiated by the reported findings and conclusions” (General Accounting Office, p. 169-170).

The KEC, under the conclusions and implications section, provides guidance for possible recommendations and explanations. This checkpoint makes a distinction between “operational recommendations” which are “those concerning the internal workings of program management and equipment choices/use” and “macro-recommendations” which are “about the disposition of the whole program” (Scriven, 2006, p. 12). The former are often easily and appropriately included in evaluations while the later should only be included if the evaluator has extensive knowledge of the program decision-makers and “knowledge of all the internal . . . management options available and the probable outcomes if they are implemented” (p. 12). LPE reports typically include operational recommendations but often macro-recommendations are also included. LPE staff members certainly possess extensive knowledge of program decision-makers (i.e. legislators). Their knowledge of internal management options is less certain.

These standards were translated to the metaevaluation criteria as follows.

Excellent: All recommendations presented are specific, measurable, and logically follow findings and conclusions.

Good: Most recommendations presented are specific, measurable, and logically follow findings and conclusions.

Fair: Recommendations logically follow findings and conclusions but are unspecific or not measurable.

Poor: Recommendations do not logically follow findings and conclusions.

Timely

Timeliness is linked to utility by both the GAGAS and the PES. The GAGAS standard for timeliness (8.39) provides that “[t]o be of maximum use, the audit report needs to provide relevant information in time to respond to officials of the audited entity, legislative officials, and other users’ legitimate needs” (General Accounting Office, p. 173). The subsequent standard (8.40) suggests that auditors consider interim reporting on “significant matters” (p. 173).

The PES utility substandard (U6) concerning report timeliness and dissemination also stresses the need to issue “[s]ignificant interim findings and evaluation reports” to “intended users” in a “timely fashion” (Joint Committee, p. 53).

These standards were translated to the metaevaluation criteria as follows.

Excellent: The report was completed within the requirement (if any) of the authorization.

Good: The report’s time requirement was extended by the report’s authorizer and the report was completed within this extended deadline.

Fair: The report was not completed within the requirements but legitimate reasons are given for this.

Poor: The report was not completed with the requirements with no reasonable explanation.

CHAPTER IV

FINDINGS

This chapter presents findings from the metaevaluation of the legislative program evaluation reports sample, the metaevaluation of the non-legislative program evaluation reports sample, and the analysis of the recommendations presented in the legislative program evaluation report sample, the Louisiana legislative program evaluation reports, and the non-legislative program evaluation reports sample. An additional section addresses the comparisons criterion, as that criterion showed the greatest variability of grades for the sample LPE reports.

LPE Metaevaluations

Overall, the LPE reports fared very well against the metaevaluation criteria. Given that the criteria were drawn largely from both the GAGAS and PES this is not surprising. Fifty-five percent of reports specified that they were conducted in accordance with the GAGAS.⁵ An additional eight percent cited unspecified governmental auditing standards. While four percent cited “applicable evaluation standards” no reports specifically cited the PES.⁶ The remaining thirty-seven percent of the reports did not cite any standards.

⁵ One of these reports, from Illinois, cited both the GAGAS and Illinois Auditor General standards.

⁶ All of these reports were from Florida’s Office of Program Policy Analysis and Government Accountability (OPPAGA). OPPAGA’s website cites the PES as the standards followed for research conducted by their office (see <http://www.oppaga.state.fl.us/about.html>).

Table 8. LPE Reports Sample: Standards Cited.

| Report type | GAGAS (%) | Unspecified Standards (%) | None (%) |
|------------------------------|-----------|---------------------------|-----------|
| Other (23 total) | 8 (34.8) | 0 (0.0) | 15 (65.2) |
| Evaluation (25 total) | 3 (12.0) | 5 (20.0) | 17 (68.0) |
| Performance audit (47 total) | 43 (91.5) | 2 (4.3) | 2 (4.3) |
| Sunset (5 total) | 1 (20.0) | 1 (20.0) | 3 (60.0) |

Table 8 shows the breakdown of the standards cited in sample LPE reports by classification of each report (other, evaluation, performance audit, or sunset). As one might expect, a high percentage, 91.5 percent, of the performance audits cited the GAGAS. Almost thirty-five percent of reports classified as other also cited the GAGAS while a small percentage, 12 percent, of evaluations cited the GAGAS.

Preliminaries

The preliminaries criteria saw general agreement across the four report categories—other, evaluation, performance audit, and sunset. Majorities of all report types were graded excellent on the scope and objectives and stakeholders criteria, fair on the values criterion, and poor on the quality control criterion.

A strong majority, 78.7 percent, of performance audits were graded excellent on the standards identification criterion. Over two-thirds of evaluations were graded poor on this criterion. A fifth of evaluations partially met this criterion by referencing general standards but not specifying which standards. Slightly less than five percent of performance audits referenced unspecified standards. Table 9 shows the ratings for all three report types on the preliminaries criteria.

Table 9. LPE Metaevaluation Ratings: Preliminaries.

| Report type | S&O (%) | Values (%) | QC (%) | Standards (%) | Stakeholders (%) |
|-------------------|-----------|------------|-----------|---------------|------------------|
| Excellent | | | | | |
| Other | 19 (82.6) | 0 (0.0) | 0 (0.0) | 5 (21.7) | 17 (73.9) |
| Evaluation | 22 (88.0) | 0 (0.0) | 0 (0.0) | 3 (12.0) | 21 (84.0) |
| Performance audit | 26 (55.3) | 1 (2.1) | 0 (0.0) | 37 (78.7) | 40 (85.1) |
| Sunset | 3 (60.0) | 1 (20.0) | 0 (0.0) | 0 (0.0) | 3 (60.0) |
| Total | 70 (70.0) | 2 (2.0%) | 0 (0.0%) | 45 (45.0%) | 81 (81.0) |
| Good | | | | | |
| Other | 3 (13.0) | 2 (8.7) | 0 (0.0) | 3 (13.0) | 2 (8.7) |
| Evaluation | 2 (8.0) | 0 (0.0) | 2 (8.0) | 0 (0.0) | 2 (8.0) |
| Performance audit | 20 (42.6) | 0 (0.0) | 8 (17.0) | 6 (12.8) | 3 (6.4) |
| Sunset | 1 (20.0) | 0 (0.0) | 0 (0.0) | 1 (20.0) | 0 (0.0) |
| Total | 26 (26.0) | 2 (2.0) | 10 (10.0) | 10 (10.0) | 7 (7.0) |
| Fair | | | | | |
| Other | 0 (0.0) | 20 (87.0) | 1 (4.3) | 0 (0.0) | 3 (13.0) |
| Evaluation | 1 (4.0) | 24 (96.0) | 0 (0.0) | 5 (20.0) | 2 (8.0) |
| Performance audit | 0 (0.0) | 46 (97.9) | 2 (4.3) | 2 (4.3) | 4 (8.5) |
| Sunset | 1 (20.0) | 4 (80.0) | 0 (0.0) | 1 (20.0) | 2 (40.0) |
| Total | 2 (2.0) | 94 (94.0) | 3 (3.0) | 8 (8.0) | 11 (11.0) |
| Poor | | | | | |
| Other | 1 (4.3) | 1 (4.3) | 22 (95.7) | 15 (65.2) | 1 (4.3) |
| Evaluation | 0 (0.0) | 1 (4.0) | 23 (92.0) | 17 (68.0) | 0 (0.0) |
| Performance audit | 1 (2.1) | 0 (0.0) | 37 (78.7) | 2 (4.3) | 0 (0.0) |
| Sunset | 0 (0.0) | 0 (0.0) | 5 (100.0) | 3 (60.0) | 0 (0.0) |
| Total | 2 (2.0) | 2 (2.0) | 87 (87.0) | 37 (37.0) | 1 (1.0) |
| Not applicable | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Evaluation | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Sunset | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

Note. S&O = scope and objectives; QC = quality control.

Methodology

Three criteria under the methodology section—complete, defensible and rights of human subjects—were overwhelmingly graded excellent or good (see Table 10).

Table 10. LPE Metaevaluation Ratings: Methodology.

| Report type | Complete (%) | Accurate (%) | Defensible (%) | Rights (%) | Compare (%) |
|-------------------|--------------|--------------|----------------|-------------|-------------|
| Excellent | | | | | |
| Other | 22 (95.7) | 12 (52.2) | 23 (100.0) | 0 (0.0) | 3 (13.0) |
| Evaluation | 24 (92.0) | 16 (64.0) | 25 (100.0) | 0 (0.0) | 11 (44.0) |
| Performance audit | 46 (97.9) | 36 (76.6) | 46 (97.9) | 0 (0.0) | 6 (12.8) |
| Sunset | 4 (80.0) | 1 (20.0) | 5 (100.0) | 0 (0.0) | 1 (20.0) |
| Total | 95 (95.0) | 65 (65.0) | 99 (99.0) | 0 (0.0) | 21 (21.0) |
| Good | | | | | |
| Other | 0 (0.0) | 2 (8.7) | 0 (0.0) | 23 (100.0) | 0 (0.0) |
| Evaluation | 1 (4.0) | 1 (4.0) | 0 (0.0) | 25 (100.0) | 2 (8.0) |
| Performance audit | 0 (0.0) | 6 (12.8) | 0 (0.0) | 47 (100.0) | 2 (4.3) |
| Sunset | 0 (0.0) | 1 (20.0) | 0 (0.0) | 5 (100.0) | 2 (40.0) |
| Total | 1 (1.0) | 10 (10.0) | 0 (0.0) | 100 (100.0) | 6 (6.0) |
| Fair | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 5 (21.7) |
| Evaluation | 1 (4.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 12 (48.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 21 (44.7) |
| Sunset | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (20.0) |
| Total | 1 (1.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 36 (36.0) |
| Poor | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 15 (65.2) |
| Evaluation | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 18 (38.3) |
| Sunset | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (20.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 37 (37.0) |
| Not applicable | | | | | |
| Other | 1 (4.3) | 9 (39.1) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Evaluation | 0 (0.0) | 8 (32.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 1 (2.1) | 5 (10.6) | 1 (2.1) | 0 (0.0) | 0 (0.0) |
| Sunset | 1 (20.0) | 3 (60.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 3 (3.0) | 25 (25.0) | 1 (1.0) | 0 (0.0) | 0 (0.0) |

Note. Rights = rights of human subjects; compare = comparisons.

Large majorities of all report types were graded excellent or good on the accuracy criterion. A fourth of all reports were judged not applicable on this criterion due to the reports lack of comments from the evaluated agency or department. Many of these reports were of the variety that would not lend them to including agency or department comments (such as best practice reviews or evaluations of overarching policy areas).

The comparisons criterion saw the reports fairly evenly distributed with about a third scoring excellent or good (27 percent), about a third fair (36 percent), and about a third poor (37 percent). Over half of the evaluations, 52 percent, were graded excellent or good on this criterion while no evaluations were graded poor. In comparison, less than a fifth, 17.1 percent, of performance audits were graded excellent or good on comparisons while more than a third, 38.3 percent, were graded poor.

Results

The context, clarity, concision, findings and conclusions, and recommendations criteria all were graded almost universally excellent or good (see Table 11). Over ninety percent of all reports were not able to be evaluated on the timely criterion because information regarding requirements for completing the report was not presented. All reports that did include enough information to evaluate timeliness were graded excellent.

Discussion

Overall, the metaevaluation results show strong adherence to the criteria. Five of the six criteria in the results category (context, clear, concise, findings and conclusions, and recommendations), three of the five criteria in the methodology category (complete, defensible, and rights), and one of the criteria in the preliminaries category (scope and objectives) saw over 90 percent of reports graded as excellent or good.

Table 11. LPE Metaevaluation Ratings: Results.

| Report type | Context (%) | Clear (%) | Concise (%) | F&C (%) | Rec. (%) | Time (%) |
|-------------------|-------------|------------|-------------|------------|-----------|------------|
| Excellent | | | | | | |
| Other | 18 (78.3) | 21 (91.3) | 23 (100.0) | 23 (100.0) | 21 (91.3) | 2 (8.7) |
| Evaluation | 24 (96.0) | 25 (100.0) | 25 (100.0) | 24 (96.0) | 22 (88.0) | 0 (0.0) |
| Performance audit | 36 (76.6) | 43 (91.5) | 47 (100.0) | 45 (95.7) | 42 (89.4) | 3 (6.4) |
| Sunset | 4 (80.0) | 5 (100.0) | 2 (40.0) | 5 (100.0) | 5 (100.0) | 4 (80.0) |
| Total | 82 (82.0) | 94 (94.0) | 97 (97.0) | 97 (97.0) | 90 (90.0) | 9 (9.0) |
| Good | | | | | | |
| Other | 1 (4.3) | 2 (8.7) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Evaluation | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 10 (21.3) | 4 (8.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Sunset | 1 (20.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 12 (12.0) | 2 (2.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Fair | | | | | | |
| Other | 2 (8.7) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Evaluation | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 1 (2.1) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Sunset | 0 (0.0) | 0 (0.0) | 3 (60.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 4 (4.0) | 0 (0.0) | 3 (3.0) | 1 (1.0) | 0 (0.0) | 0 (0.0) |
| Poor | | | | | | |
| Other | 2 (8.7) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Evaluation | 1 (4.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (2.1) | 0 (0.0) | 0 (0.0) |
| Sunset | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 2 (2.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Not applicable | | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (8.7) | 21 (91.3) |
| Evaluation | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (4.0) | 3 (12.0) | 25 (100.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (2.1) | 5 (10.6) | 44 (93.6) |
| Sunset | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (20.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (2.0) | 10 (10.0) | 91 (91.0) |

Note. F&C = findings and conclusions; Rec. = recommendations; Time = timely.

Several of these criteria—clear, concise, complete, accurate and defensible—were difficult to grade. Though the reports as a group deserved high ratings, the difficulty of objectively rating these categories across reports should not be underestimated (especially for the complete, accurate and defensible criteria). The clear and concise criteria were somewhat easier to apply.

Other areas on which a high percentage of reports were graded excellent or good—scope and objectives, context, findings and conclusions, and recommendations—entailed more specific rules for application. Therefore the findings for these criteria should be given greater confidence.

The remaining high scoring criteria—rights of human subjects (100 percent of reports graded good) and the accurate criterion (on which 75 percent of reports were graded excellent or good with the remaining 25 percent not applicable)—both entail rules that lend them to asymmetrical evaluation. A report met the accurate criterion if it did not contain an objection to its accuracy in the agency comments to the report. Agencies that received erroneous favorable findings would be unlikely to object. Therefore this criterion, as applied in this study, could only be credibly said to have judged if the LPE report was inaccurate in the *negative* direction, not the positive. The human subjects criterion was judged as met if a report did not include evidence of violating the rights of human subjects. This would obviously miss reports that violated human subjects' rights but did not include evidence of this violation in the report. These two criteria are especially troublesome when conducting a metaevaluation solely on the final report.

Though the time criterion is very important for judging the value of LPE reports, the lack of information provided in the actual text of the reports did not allow for an accurate review for 91 percent of the reports. This criterion came directly from the GAGAS performance auditing standards. It is likely that the LPE reports are being completed in a timely manner, but the reports themselves are not specifying this.

This analysis of the metaevaluations highlights a few interesting differences between performance audits and evaluations. Performance audits were much more likely

to be graded excellent or good on the standards criteria (91.5 percent) than were the evaluations (12 percent). This suggests that performance auditors are more professionally linked to the GAGAS than evaluators are linked to the PES or other evaluation standards. Evaluations were more likely to be graded excellent or good on the comparisons criteria (52 percent) than were performance audits (14.9 percent). Finally, on the quality control criteria 78.7 percent of performance audits versus 92 percent of evaluations were graded poor. This criterion was judged by the presence or absence, in a report, of internal quality control measures and any external peer reviews. Again, this may be a case of LPE units following good quality control procedures but not specifying them in their final reports.

Comparisons

As discussed above, the comparisons criterion showed the most variability for the LPE reports. The discussion below explores these differences further and also contrasts the performance of the LPE reports with the performance of the non-LPE reports on the comparisons criterion.

Among the LPE reports, evaluations fared far better on the comparisons criterion than did performance audits, sunset reports or other reports. No evaluations were graded poor, while 38.3 percent of performance audits and 65.2 percent of other reports did.

Several caveats apply to the discussion of comparisons in LPE reports. The metaevaluation criterion for comparisons only looks at what type of comparative studies were cited, not how well they were used, how closely they fit the situation in the extant evaluation, or how methodologically sound was the comparative study being cited.

Table 12. LPE Reports Using Comparisons.

| Report type | Comparisons (%) | Cites comparison in objectives (%) | Cites specific state LPE report(s) (%) | Cites national or nonspecific state report(s) (%) |
|-------------------|-----------------|------------------------------------|--|---|
| Excellent | | | | |
| Other | 3 (13.0) | 1 (33.3) | 1 (33.3) | 2 (66.7) |
| Evaluation | 11 (44.0) | 3 (27.3) | 2 (18.2) | 9 (81.8) |
| Performance audit | 6 (12.8) | 2 (33.3) | 2 (33.3) | 5 (83.3) |
| Sunset | 1 (20.0) | 0 (0.0) | 1 (100.0) | 1 (100.0) |
| Total | 21 (21.0) | 6 (28.6) | 6 (28.6) | 17 (81.0) |
| Good | | | | |
| Other | 0 (0.0) | | | |
| Evaluation | 2 (8.0) | 0 (0.0) | 0 (0.0) | 2 (100.0) |
| Performance audit | 2 (4.3) | 0 (0.0) | 0 (0.0) | 2 (100.0) |
| Sunset | 2 (40.0) | 0 (0.0) | 0 (0.0) | 2 (100.0) |
| Total | 6 (6.0) | 0 (0.0) | 0 (0.0) | 6 (100.0) |

Table 12 shows the LPE reports that were graded excellent or good on the comparisons criterion. Twenty-seven percent of LPE reports achieved these grades. The remaining seventy-three percent of reports were graded either poor (i.e., they did not include any reference to similar programs/agencies or to audits or evaluations of similar programs or agencies) or fair (the report addressed similar programs/agencies but did not address evaluations or audits of these programs/agencies). The table identifies, by report types, whether the report's objectives specifically call for a comparative examination of the program/agency, whether the report cites a specific LPE report or reports from other states, and whether the report cites national or nonspecific LPE reports. A few reports cited both specific LPE reports from other states and national reports concerning the program/agency under review.

A minority (28.6 percent) of the reports that were graded excellent contained objectives that specifically called for a comparative review. While none of the six reports scoring good called for a comparative review in their objectives. Similarly, 28.6 percent of reports scoring excellent cited specific LPE reports (evaluations, audits, etc.) from

other states. No reports that were graded good cited specific LPE reports. The vast majority of excellent reports (81 percent) and all of the good reports cited national studies or nonspecific state studies.

Table 13 shows the ratings on the comparisons criteria for the LPE reports and the non-LPE reports. The majority of reports in each category (non-LPE and LPE) were judged fair or poor. However, 28 percent of LPE reports were graded excellent or good while only one of the ten non-LPE reports was graded good.

Table 13. Comparisons Criterion: Non-LPE Versus LPE Reports.

| | Non-LPE comparisons (%) | LPE comparisons (%) |
|-------------------|-------------------------|---------------------|
| | Excellent | |
| Other | 0 (0.0) | 3 (13.0) |
| Evaluation | | 11 (44.0) |
| Performance audit | 0 (0.0) | 6 (12.8) |
| Sunset | | 1 (20.0) |
| Total | 0 (0.0) | 21 (21.0) |
| | Good | |
| Other | 0 (0.0) | 0 (0.0) |
| Evaluation | | 2 (8.0) |
| Performance audit | 1 (12.5) | 2 (4.3) |
| Sunset | | 2 (40.0) |
| Total | 1 (10.0) | 6 (6.0) |
| | Fair | |
| Other | 1 (50.0) | 5 (21.7) |
| Evaluation | | 12 (48.0) |
| Performance audit | 5 (62.5) | 21 (44.7) |
| Sunset | | 1 (20.0) |
| Total | 6 (60.0) | 36 (36.0) |
| | Poor | |
| Other | 1 (50.0) | 15 (65.2) |
| Evaluation | | 0 (0.0) |
| Performance audit | 2 (25.0) | 18 (38.3) |
| Sunset | | 1 (20.0) |
| Total | 3 (30.0) | 36 (36.0) |
| | Not applicable | |
| Other | 0 (0.0) | 0 (0.0) |
| Evaluation | | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) |
| Sunset | | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) |

Legislative Versus Non-legislative Program Evaluation Units

This section examines legislatively based program evaluation units and similar units located in the executive branch of state governments. Non-legislative program evaluation units that exist in the ten states that did not have a LPE unit issuing reports during 2001 through 2005 were identified (Alabama, Delaware, Iowa, Maine, New Jersey, North Carolina, North Dakota, Oklahoma, South Dakota, and Vermont). Four of these states (see Table 3, p. 87) had an executive branch agency issuing performance audit or evaluation reports during 2001 through 2005.

These four units issued a total of 53 reports from 2001 through 2005. These 53 reports were classified as either performance audits or other reports (see Table 4, p. 88). A sample of 10 non-LPE reports was selected using the same procedure described to select the LPE report sample. The sample of non-LPE reports was metaevaluated using the same criteria applied to the NPE report sample. Tables 14, 15, and 16 show the metaevaluation ratings for the non-LPE report sample.

Preliminaries

The preliminaries criteria saw general agreement across the two report categories—other and performance audit—although the small number of other reports makes across category comparisons difficult. Majorities of all non-LPE reports were graded excellent on the scope and objectives, standards, and stakeholder criteria, fair on the values criterion, and poor on the quality control criterion. These results are similar to

the results for the LPE reports. The only difference was seen on the standards criteria. While 90 percent of the non-LPE reports were graded excellent on this criterion, only 45 percent of LPE reports were graded excellent (see Table 9, p. 117). Table 14 shows the ratings for both report types on the preliminaries criteria.

Table 14. Non-LPE Metaevaluation Ratings: Preliminaries.

| | S&O (%) | Values (%) | QC (%) | Standards (%) | Stakeholders (%) |
|-------------------|-----------|------------|------------|---------------|------------------|
| Excellent | | | | | |
| Other | 1 (50.0) | 0 (0.0) | 0 (0.0) | 1 (50.0) | 2 (100.0) |
| Performance audit | 7 (87.5) | 0 (0.0) | 0 (0.0) | 8 (100.0) | 8 (100.0) |
| Total | 8 (80.0) | 0 (0.0) | 0 (0.0) | 9 (90.0) | 10 (100.0) |
| Good | | | | | |
| Other | 1 (50.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 1 (12.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 20 (20.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Fair | | | | | |
| Other | 0 (0.0) | 2 (100.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 8 (100.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 10 (100.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Poor | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 2 (100.0) | 1 (50.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 8 (100.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) | 10 (100.0) | 1 (10.0) | 0 (0.0) |
| Not applicable | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

Note. S&O = scope and objectives; QC = quality control.

Methodology

Table 15 presents the results for non-LPE reports on the methodology criteria. Three criteria under the methodology section—complete, accurate, and defensible—were uniformly graded excellent.

The two reports classified as other were graded good on the rights criterion while the eight performance audits were graded fair on that criterion.

The comparisons criterion saw the reports distributed among the good, fair, and poor ratings. One other report was graded fair and one poor. Five of the eight performance audits were graded fair, one graded good, and the remaining two graded poor.

There were virtually no differences between the LPE and non-LPE reports on the complete and defensible criteria. Both were graded very highly on each. On the accurate criterion, the non-LPE reports all were graded excellent while the 65 percent of the LPE reports were graded excellent. A quarter of the LPE reports were not graded on the accuracy criterion because they did not include comments from the agency/office under review. All of the non-LPE reports included agency comments.

On the comparisons criterion, 22 percent of LPE reports were graded excellent while no non-LPE reports were graded excellent. This difference may be explained in part by the absence of evaluations among the non-LPE reports. Forty-four percent of the LPE evaluations were graded excellent.

Results

Ratings on the results criteria were very similar for non-LPE and LPE reports. The context, clarity, concision, findings and conclusions, and recommendations criteria all were graded almost universally excellent or good. Almost all non-LPE and LPE reports were not able to be evaluated on the timely criterion because information

Table 15. Non-LPE Metaevaluation Ratings: Methodology.

| | Complete (%) | Accurate (%) | Defensible (%) | Rights (%) | Compare (%) |
|-------------------|--------------|--------------|----------------|------------|-------------|
| Excellent | | | | | |
| Other | 2 (100.0) | 2 (100.0) | 2 (100.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 8 (100.0) | 8 (100.0) | 8 (100.0) | 0 (0.0) | 0 (0.0) |
| Total | 10 (100.0) | 10 (100.0) | 10 (100.0) | 0 (0.0) | 0 (0.0) |
| Good | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (100.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (12.5) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (20.0) | 1 (10.0) |
| Fair | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (50.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 8 (100.0) | 5 (62.5) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 8 (80.0) | 6 (60.0) |
| Poor | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (50.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (25.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 3 (30.0) |
| Not applicable | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

Note. Compare = comparisons.

regarding requirements for completing the report was not presented. The single report that did include enough information to evaluate timeliness was graded excellent. Table 16 shows ratings for the non-LPE reports on the results criteria.

Table 17 shows the breakdown of the standards cited in sample non-LPE reports by classification of each report (other, evaluation, performance audit, or sunset). Nine out of the ten performance audits cited the GAGAS. One of two reports classified other also cited the GAGAS. The remaining report did not cite any standards. None of the sample, or population, non-LPE reports were evaluations or sunset reports.

Table 16. Non-LPE Metaevaluation Ratings: Results.

| | Context (%) | Clear (%) | Concise (%) | F&C (%) | Rec. (%) | Time (%) |
|-------------------|-------------|------------|-------------|------------|------------|-----------|
| Excellent | | | | | | |
| Other | 1 (50.0) | 2 (100.0) | 2 (100.0) | 2 (100.0) | 2 (100.0) | 0 (0.0) |
| Performance audit | 6 (75.0) | 8 (100.0) | 8 (100.0) | 8 (100.0) | 8 (100.0) | 1 (12.5) |
| Total | 7 (70.0) | 10 (100.0) | 10 (100.0) | 10 (100.0) | 10 (100.0) | 1 (10.0) |
| Good | | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Fair | | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Poor | | | | | | |
| Other | 1 (50.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit | 2 (20.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Total | 3 (30.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Not applicable | | | | | | |
| Other | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (100.0) |
| Performance audit | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 7 (87.5) |
| Total | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 9 (90.0) |

Note. F&C = findings and conclusions; Rec. = recommendations; Time = timeliness.

Table 17. LPE Reports Sample: Standards Cited.

| Report type | GAGAS (%) | Unspecified standards (%) | None (%) |
|-----------------------------|-----------|---------------------------|----------|
| Other (2 total) | 1 (50.0) | 0 (0.0) | 1 (50.0) |
| Evaluation (0 total) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Performance audit (8 total) | 8 (100.0) | 0 (0.0) | 0 (0.0) |
| Sunset (0 total) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

Report Recommendations

Table 18 shows information concerning recommendations in the LPE reports. As a whole, the LPE reports are more likely to contain recommendations for the agency under review (75 percent of reports) than for the legislature (43 percent of reports). This

difference was reflected in all report types except for sunset reports. This should be expected because sunset reports are specifically designed to report to the legislature on whether, and in what form, an agency should be reauthorized. Conversely, performance audits were the report type most likely to contain agency recommendations (85.1 percent of reports) and least likely to contain legislative recommendations (29.8 percent of reports). This follows from the GAGAS requirements which do not address legislative recommendations but deal extensively with agency recommendations.

Table 18. LPE Reports with Recommendations.

| Report type | Reports (%) | Reports (%) | Rec. per report | Reports (%) | Rec. per report | Reports (%) | Rec. per report |
|-------------------|-------------|-------------|-----------------|-------------|-----------------|-------------|-----------------|
| | No rec. | Agency rec. | | Leg. rec. | | Both | |
| Other | 2 (8.7) | 16 (69.9) | 4.6 | 14 (60.9) | 1.9 | 9 (39.1) | 7.4 |
| Evaluation | 6 (24.0) | 17 (68.0) | 6.8 | 11 (44.0) | 1.8 | 9 (36.0) | 8.2 |
| Performance audit | 6 (12.8) | 40 (85.1) | 9.6 | 14 (29.8) | 2.9 | 13 (27.7) | 12.4 |
| Sunset | 0 (0.0) | 2 (40.0) | 2.5 | 4 (80.0) | 14.0 | 1 (20.0) | 26.0 |
| Total | 14 (14.0) | 75 (75.0) | 7.7 | 43 (43.0) | 3.3 | 32 (32.0) | 10.2 |

Note. Rec. = recommendations; leg. = legislative.

Table 19 concerns recommendations in non-LPE reports. As with the LPE reports, these reports are much more likely to contain agency recommendations (92.5 percent of reports) than to contain legislative recommendations (49.1 percent of reports). Among the non-LPE reports, performance audits (57.1 percent of reports) are more likely than other reports (33.3 percent of reports) to contain legislative recommendations.

All Louisiana reports are performance audits. Almost all these reports contain agency recommendations (95.2 percent of reports) and two-thirds contain legislative recommendations. Only 4.8 percent of reports contain no recommendations.

Table 19. Non-LPE Reports with Recommendations.

| Report type | Reports (%) | Reports (%) | Rec. per report | Reports (%) | Rec. per report | Reports (%) | Rec. per report |
|-------------------|-------------|-------------|-----------------|-------------|-----------------|-------------------|-----------------|
| | No rec. | Agency rec. | | Leg. rec. | | Agency & leg rec. | |
| Other | 1 (5.5) | 17 (94.4) | 10.5 | 6 (33.3) | 2.3 | 6 (33.3) | 14.5 |
| Performance audit | 2 (5.7) | 32 (91.4) | 13.9 | 20 (57.1) | 2.2 | 19 (54.3) | 16.4 |
| Total | 3 (5.7) | 49 (92.5) | 12.7 | 26 (49.1) | 2.2 | 25 (47.2) | 16.0 |

Note. Rec. = recommendations; leg. = legislative.

Interestingly, 14 percent of LPE reports contained a recommendation to eliminate an agency or program while no non-LPE reports and no Louisiana LPE reports contained such a recommendation. This situation is certainly influenced by the presence of sunset reports in the LPE sample while the non-LPE reports and Louisiana reports do not include sunset reports. However, three of the fourteen performance audit LPE reports (21.4 percent) also contain a recommendation for elimination.

Recommendations for consolidation of programs and agencies are slightly more common among LPE reports than non-LPE and Louisiana reports. Almost all types of reports contained at least one report with a consolidation recommendation. The exception was the category of other non-LPE reports.

Non-LPE reports (30.8 and 23.1 percent) and Louisiana reports (21.4 and 42.9 percent) were slightly more likely than LPE reports (16.3 and 20.9 percent) to recommend further study or a clarification of the law.

Comparisons between LPE reports and non-LPE reports disclosed no great differences. Large majorities of all reports from these three groups contain recommendations. LPE reports were only slightly more likely to address their recommendations to the legislature than were non-LPE reports. There was also a small difference in the frequency of recommendations for program or agency elimination. LPE

were more likely than non-LPE reports to recommend elimination. One might surmise that one executive branch agency would be unlikely to recommend the elimination of another executive branch agency. Legislative agencies may have less hesitation in this regard.

Table 20 shows the types of legislative recommendations for LPE reports, non-LPE reports, and Louisiana LPE reports. The number listed for each category indicates the number of reports that contained a recommendation of the type specified. Reports could contain recommendations of several different types.

Table 20. LPE, Non-LPE, and Louisiana Reports: Recommendations.

| Report type | Further study (%) | Clarify (%) | Substantive change (%) | Eliminate (%) | Consolidate (%) |
|-----------------------|-------------------|-------------|------------------------|---------------|-----------------|
| LPE reports | | | | | |
| Other | 3 (21.4) | 1 (7.1) | 12 (85.7) | 0 (0.0) | 1 (7.1) |
| Evaluation | 2 (18.2) | 1 (9.1) | 10 (90.9) | 0 (0.0) | 1 (9.1) |
| Performance audit | 1 (7.1) | 5 (35.7) | 13 (92.9) | 3 (21.4) | 2 (14.3) |
| Sunset | 1 (25.0) | 2 (50.0) | 4 (100.0) | 3 (75.0) | 3 (75.0) |
| Total | 7 (16.3) | 9 (20.9) | 39 (90.7) | 6 (14.0) | 7 (16.3) |
| Non-LPE reports | | | | | |
| Other | 1 (16.7) | 2 (33.3) | 5 (83.3) | 0 (0.0) | 0 (0.0) |
| Performance audit | 7 (35.0) | 4 (20.0) | 17 (85.0) | 0 (0.0) | 3 (15.0) |
| Total | 8 (30.8) | 6 (23.1) | 22 (84.6) | 0 (0.0) | 3 (11.5) |
| Louisiana LPE reports | | | | | |
| Performance audit | 6 (21.4) | 12 (42.9) | 23 (82.1) | 0 (0.0) | 3 (10.7) |

One prominent difference between LPE and non-LPE reports was the absence of evaluations among the non-LPE reports. Perhaps as a consequence, the non-LPE reports were more likely to cite the GAGAS as their governing standards. The small population of reports and states with non-LPE units makes it difficult to draw wider conclusions from these comparisons with the LPE reports.

CHAPTER V

CONCLUSIONS

This study set out to describe how U.S. state legislative staffs conduct evaluations. The study addressed the ubiquity of state LPE units, the standards those units follow, the recommendations that LPE reports proffer, and the quality of the reports on several criteria. In the process of this investigation there was an examination of the feasibility of using metaevaluation to evaluate a large number of reports using solely the information contained in the reports themselves.

LPE Units and Reports

Forty of the fifty U.S. state legislatures had a LPE unit that published at least one legislative program evaluation report during the five-year period covered in the study. One state, Texas, had two such units (see Appendix B). Four of the ten states without LPE units within their respective legislatures had executive branch evaluation or performance audit units (see Table 3, p. 87).

State LPE units, and consequently the reports they produce, are overwhelmingly more connected to performance auditing and the GAGAS than to evaluation and evaluation standards (whether the Joint Committee Standards or the AEA's *Guiding Principles for Evaluators*). Approximately half of LPE reports are identified as performance audits while only a quarter are identified as evaluations (Table 4, p. 88).

Over half the reports cited the GAGAS as their governing standards while none specifically cited the PES or the AEA's *Guiding Principles* (Table 8, p. 116).

This heavy reliance on the GAGAS can be explained due to the emergence of state legislative program evaluation largely from the financial auditing units already existing in state legislatures. Also, the GAGAS are the official standards for government auditing while the PES are not. Still, given the influence of the field of evaluation on legislative program evaluation as seen by the choice of the name for the premier professional group of these state units—the National Legislative Program Evaluation Society—one would expect more explicit reference to evaluation standards.

State LPE reports direct most of their recommendations toward the agencies or office they are evaluating instead of toward the state's legislators (see Table 18, p. 131). Performance audits are slightly more likely than evaluations to offer recommendations to agencies while evaluations are slightly more likely than performance audits to direct their recommendations to the legislature. The types of recommendations offered by performance audits and evaluations did not differ much though performance audits were slightly more likely to recommend program or agency elimination or consolidation.

Applying the metaevaluation criteria to the LPE reports resulted in consistently high grades, across all four report categories, on most criteria. Two criteria on which performance audits and evaluations differed markedly were standards cited and comparisons.

Performance audits were highly likely to cite the standards under which the report was produced while evaluations very rarely cited applicable standards (see Table 8, p. 116). Like many differences between performance audits and evaluations this may be a

result of the auditing culture that took root in financial audits and migrated to performance audits. The standards cited were predominately the GAGAS with a low percentage of reports citing unspecified standards.

The use of comparisons in the LPE reports was an interesting finding that saw a great deal of variation among the different report types and even among different reports of the same type. Roughly a third of all LPE reports were graded excellent or good on the comparisons criterion. Another third of reports were graded fair. The final third was graded poor—reflecting no mention of comparisons in the report (see Table 10, p. 118). Evaluations were more likely to be graded excellent or good on this criterion than were performance audits.

The examination of comparisons should be of paramount importance to legislative program evaluation. Legislatures allocate resources among the competing needs within their jurisdictions. They need rigorous examination of the choices they *have* made, and the choices they *could* make. LPE units could provide much more of this type of information to their respective legislatures. In addition to providing evaluations focused on the specific goals and objectives of each program or policy under review, LPE units could expand their reviews to include examinations of different ways to meet the same needs.

Some state LPE units accomplish this through best practice reviews or policy briefs. These reports usually do not focus on a particular, existing program or policy; rather they look at a problem or need and examine different options for addressing it. The concept of examining best practices could be incorporated into performance audit and evaluation reports on a smaller scale. Each performance audit or evaluation could include

a section highlighting best practices and comparing the program or policy under review to the identified best practices.

Such a step could aid legislatures in modifying and designing policy. It would also facilitate the transfer of successful policies and programs from one jurisdiction to another as each state would constantly be examining how other states address similar needs. A comparative or best practices approach would most likely need to be proposed by a state legislature, through statute, or by a legislative committee overseeing an LPE unit. LPE units currently adhere closely to legislative direction when pursuing their reports.

The LPE reports, especially the performance audits, are very goal-directed. The objectives for the report are laid out clearly—as required by the GAGAS—and the reports seldom stray beyond addressing topic suggested when the report is requested. Among reports graded good or excellent on the comparisons criterion report objectives seldom contained specific instructions to compare the agency or program under review to either a different approach from another state or to a different idea for addressing the same needs the program or policy under review was designed to address (see Table 13, p. 125). Several reports ventured beyond the specific call of their objectives and investigated comparisons. If more organizations that, and people who, commission these reports (legislative commissions, legislative committees, and LPE unit leaders) specifically call for comparative explorations these explorations will certainly increase.

The format of LPE reports, while varying across the different states and units, is similar enough that it leads to greater ease of comparing reports than would likely be the case trying to compare evaluation reports in general. This consistency of style is likely a

product of the GAGAS, given its wide use in the LPE reports sample. The PES also contain a standard (the program documentation standard—A1) that stresses the importance of program documentation for enabling comparisons with other programs. So there is some recognition of the importance of comparisons in the GAGAS and the PES this is an area where each could borrow from Scriven's KEC.

Metaevaluation

In addition to exploring the specific area of state legislative program evaluation reports, this study also sought to test a methodological model—that of using metaevaluation to examine a large number of reports. The results from this attempt are mixed. While the use of metaevaluation in this way can lead to greater insight on what specific aspects of evaluation reports are excelling or failing, accurately and fairly evaluating reports of this type solely from the report itself presents some major problems.

The main difficulty with metaevaluating LPE reports may lie in the nature of the reports themselves. These reports are developed for a specific audience (state legislators). To serve this audience LPE report authors seek to keep the report relatively brief. In doing so much of the technical information that is often found in complex evaluation reports is not included. This lack of information leads to necessary caveats when conducting a metaevaluation. Table 6 (p. 93) seeks to address these caveats. Eleven of the sixteen metaevaluation criteria used in this study are judged as yielding moderate confidence in the grades assigned.

A more streamlined set of metaevaluation criteria could be developed that combines several of the high scoring areas from this study. Criteria such as clear, concise, complete and defensible are certainly important for LPE reports but given the uniform success of reports in meeting these criteria they could possibly be combined. A more rigorous set of metaevaluation criteria could be applied to areas such as quality control, values, stakeholders, and timeliness if there was more agreement in the LPE field about what specific information should be included in reports.

Time constraints require that metaevaluations across a large number of evaluations be conducted using solely the evaluation reports. This leads to deficiencies on judging the quality of an evaluation on some criteria. It renders some criteria nearly impossible to judge. However, within these restraints a credible metaevaluation can be conducted and possibly contribute both the field of evaluation and to state government knowledge across many issue areas

Any evaluation is limited in the scope of information used to form evaluative conclusions. At the most basic level, no single evaluator, or team of evaluators, could possess complete information. More practically, evaluation budgets limit the resources available to collect and analyze information. Evaluation timelines also limit the extent evaluators can collect and evaluate information. Time constraints are particularly pertinent to the field of legislative program evaluation.

Metaevaluation reflects these same limits. Further limiting the basis for a metaevaluation to the reports produced by an evaluation creates still more problems. The key question is not are there limitations—there surely are—but do the limitations irreparably compromise the value of the metaevaluation.

In one sense report metaevaluations are simply evaluating the report itself and not essentially the evaluation per se. Some of the LPE metaevaluation criteria concentrate on solely the report—clarity, concision, timeliness. Other criteria straddle the issues of the essential evaluation and the reporting of the evaluation—rights of human subjects, findings and conclusions. An evaluation may competently reach findings and conclusions but not adequately report them. Conversely, the evaluation report may present convincing findings but lack the actual factual basis for such findings (these situations can range from the extreme example of fraud on the part of the evaluators to more mundane instances of errors of interpretation or errors of significance).

Metaevaluation can be a useful tool for discovering the characteristics of a general class of evaluation reports. This study has determined that state LPE reports are very goal oriented, rarely question the usefulness of the agencies being evaluated, conform closely to the GAGAS, and vary greatly in their use of comparisons. The metaevaluation criteria have aided in examining these reports and presenting these results. Metaevaluation across a large number of reports may be best suited to, and most useful for, academic research in specific areas, a peer review process for LPE units or other evaluation-related groups, studying policy transfer across jurisdictions (see Dolowitz & Marsh, 2000), and committees or other groups charged with creating or revising standards of practice in evaluation.

Key Criteria for Metaevaluating LPE Reports

The 16 metaevaluation criteria used in this study represent only a partial list of ideal LPE metaevaluation criteria. The addition of two criteria—metaevaluation and cost—would create a list of 18 key criteria for evaluating LPE reports.

Metaevaluation is particularly important for any evaluative endeavor. State legislatures commission LPE units to evaluate the work of others—work that is accomplished via policies, programs, or agencies. The GAGAS contends that work of performance auditors can “lead to improved government management, decision making, and oversight” (General Accounting Office, 2003, p. 5). Consequently, a review of the work of performance auditors and evaluators is necessary to improve performance audits and evaluations. The PES include a standard for metaevaluation. According to the Joint Committee metaevaluation “can help program evaluators avoid critical mistakes” (Joint Committee, 1994, p. 185). The GAGAS includes a requirement that audit organizations undergo external peer reviews at least once every three years (General Accounting Office, 2003, p. 57). These peer review include reviews of individual audits. A system that requires a metaevaluation be conducted for each audit or evaluation would greatly improve upon this periodic peer review requirement.

The cost effectiveness of an evaluation is covered clearly by the PES. However, the GAGAS includes no corresponding standard relating to the cost of a performance audit (see Chapter 2, pp. 68-69). This study did not include the cost of the LPE reports as a metaevaluation criterion due to the lack of information on costs included in the reports. However, any complete list of metaevaluation criteria for LPE reports should include the

costs incurred in producing each report. These costs should be measured against the benefits produced from the LPE report. Reports that produce greater benefits through the identification of program efficiencies, the elimination of inefficient or counterproductive programs, or the documentation of positive or negative side-effects from programs or policies under review would be a particularly good investment of LPE resources. Programs or policies that expend more public resources or have a greater scope would typically be more important to evaluate than would lower cost and more narrowly tailored programs. However, programs or policies that are narrow in scope can often produce large positive or negative side-effects. LPE reports that identify these programs and their hidden effects could be particularly cost-effective.

These 18 key metaevaluation criteria, drawn from the GAGAS for performance auditing and the PES, provide a tool for assessing the merit, worth and significance of LPE reports. Differing subject matter and practical considerations will necessitate that not all LPE report will be exemplary on all criteria. However, all LPE reports should adequately address each criterion.

Directions for Further Research

This study has presented a detailed view of what state LPE units are producing in their reports to their respective legislatures. The study has not addressed the extent to which the legislatures actually use the reports produced for them. There has been some recent investigation into this issue (VanLandingham, 2006) using surveys of legislators and other stakeholders. Other approaches that may prove promising are in-depth case

studies of specific LPE reports and their affects and documentary reviews of committee hearings and other legislative records in the years following LPE reports.

Another aspect of LPE reports that deserves in-depth research is the treatment of program or agency cost. This study did not directly assess the treatment of cost in its metaevaluation criteria. However, reading the reports disclosed that while program or agency expenditures and revenues are usually listed—sometimes extensively—the analysis of these costs is often non-existent or perfunctory. An entire study, replicating this study’s basic methodology of selecting a sample of LPE reports, could concentrate solely on the treatment of costs and cost analysis within the field of legislative program evaluation.

The use of comparisons in LPE reports could also be studied using a similar methodology. An examination of LPE reports that include comparative analysis could provide insight on the circumstances that contribute to the use of this important evaluative approach. This study provides evidence that LPE report objectives are not driving the use of comparisons in the report. A more focused review may uncover other, more subtle signals from legislators that cause LPE units to include comparative analysis in their reports.

Finally, the work of LPE units controlled by legislative bodies could be contrasted with similar work conducted by LPE units controlled by a state’s executive branch. This study briefly compared the two and found only minor differences. The National Legislative Program Evaluation Society includes member offices from both the executive branch and the legislative branch. A comprehensive approach to studying differences

would examine states that contain LPE units in both the executive and legislative branches of government and compare the work of these units.

Comparisons with Previous Literature

This study adds to the list of literature cited by Funkhouser (2005) concerning performance auditing in state governments in the United States. It also addresses the lack of empirical work looking directly at evaluations and performance audits cited by Pollitt and Summa (1996). This study found differences in the way LPE performance audits and evaluations use comparisons, conform to professional standards, and identify the source of their objectives.

The study takes a more narrow view of what constitutes legislative program evaluation than does the NLPES. NLPES surveys and membership lists (see Mohan, 1997; NLPES, 2004, 2007; VanLandingham, Kinton, Boerner, & Zelio, 2000) include many units that are based in a state's executive branch. This study limits LPE units to those that are directly under the control of the legislature.

A few previous studies have examined reports generated by state LPE units. Brown and Craft (1980) found that the state legislative audit-evaluation agency reports they reviewed "usually consulted reports from the GAO and other states' audit agencies" (p. 261). This study found that only 27 percent of LPE reports cited national or state reports on the topic under review. Brown and Craft, conducting their study when fewer state LPE units existed, analyzed reports from 15 state LPE units.

Looking at the Illinois Auditor General's office, Brown (1988) found that 70 percent of report recommendations are directed to the agency or program under review while 30 percent are directed to the legislature. Similarly, this study finds that 75 percent of LPE reports contain agency/program recommendations while 43 percent contain recommendations for the legislature.

This study follows a methodology used by several others in using metaevaluation to examine multiple reports in a particular field. Previous examinations using this methodology looked at evaluations of international development programs (Bollen, Paxon, & Morishima, 2005; Chatterji, 2005; Forss & Carlsson, 1997; Kruse, 2005; Leeuw & Cooksy, 2005), international agricultural programs (Cooksy & Caracelli, 2005), sports policy in the Netherlands (Hoogerwerf, 1992), after school programs (Scott-Little, Hamann, & Jurs, 2002), and criminal justice programs (Larson, et al., 1979). One previous study (Bickman, 1997) used metaevaluation to examine all evaluations, covering a variety of topics, published in two academic journals over a ten year period. This study expands the previous list to include the examination of LPE reports.

REFERENCES

- Adams, B., & Sherman, B. (1978). Sunset implementation: A positive partnership to make government work. *Public Administration Review*, 38(1), 78-81.
- Barzelay, M. (1997). Central audit institutions and performance auditing: A comparative analysis of organizational strategies in the OECD. *Governance: An International Journal of Policy and Administration*, 10(3), 235-260.
- Berry, F. S., Turcotte, J. W., & Latham, S. C. (2002). Program evaluation in state legislatures: Professional services delivered in a complex, competitive environment. *New Directions for Evaluation*, 95, 73-88.
- Bezruki, D., Mueller, J., & McKim, K. (1999). Legislative utilization of evaluations. *New Directions for Evaluation*, 81, 11-22.
- Bickman, L. (1997). Evaluating evaluation: Where do we go from here? *Evaluation Practice*, 18(1), 1-16.
- Bollen, K., Paxton, P., & Morishima, R. (2005). Assessing international evaluations: An example from USAID's democracy and governance program, *American Journal of Evaluation*, 26(2), 189-203.
- Botner, S. B. (1986, Winter). Trends and developments in state postauditing. *State and Local Government Review*, 13-19.
- Brooks, R. A. (1996). Blending two cultures: State legislative auditing and evaluation. *New Directions for Program Evaluation*, 71, 15-28.
- Brown, J. (1988). State evaluation in a legislative environment: Adapting evaluation to legislative needs. *New Directions for Program Evaluation*, 38, 49-61.
- Brown, R. E. (1976). Program evaluation audits. *The Internal Auditor*, 33(5), 85-89.
- Brown, R. E. (Ed.). (1979). *The effectiveness of legislative program review*. New Brunswick, NJ: Transaction.
- Brown, R. E., & Craft, R. (1980). Auditing and public administration: The unrealized partnership. *Public Administration Review*, 40(3), 259-265.
- Brown, R., & Pethtel, R. D. (1974). A matter of facts: State legislative performance auditing. *Public Administration Review*, 34(4), 318-327.
- Chadwin, M. L. (1974). Introduction-The nature of legislative program evaluation. In M. L. Chadwin (Ed.), *Legislative program evaluation in the states: Four case studies*

- (pp. 1-9). New Brunswick, NJ: Rutgers University, Center for Legislative Research and Service, Eagleton Institute of Politics.
- Chatterji, M. (2005). Applying the joint committee's 1994 standards in international contexts: A case study of education evaluations in Bangladesh. *Teachers College Record*, 107(10), 2372-2400.
- Chelimsky, E. (1985). Comparing and contrasting auditing and evaluation: Some notes on their relationship. *Evaluation Review*, 9(4), 483-503.
- Chelimsky, E. (1996). Auditing and evaluation: Whither the relationship? *New Directions for Program Evaluation*, 71, 61-67
- Comptroller General of the United States. (1972). *Standards for Audit of Governmental Organizations, Program, Activities, and Functions*. (GAO Publication No. 728-235/246 1-3). Washington, DC: Government Printing Office.
- Comptroller General of the United States. (1981). *Standards for audit of governmental organizations, program, activities, and functions* (GAO Publication No. 020-000-00205-1). Washington, DC: Government Printing Office.
- Comptroller General of the United States. (1988). *Government auditing standards*. (GAO Publication No. 020-000-00243-3). Washington, DC: Government Printing Office.
- 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.
- Craft, R. (1979). Legislative oversight in the states. *GAO Review*, 14(4), 22-26.
- Crane, E. (1972). Legislative Services Agencies. In R. Weber, & R. Marcelli (Eds.), *Book of the States 1972-73*. Lexington, Ky.: Council of State Governments.
- Davis, D. F. (1990). Do you want a performance audit or a program evaluation? *Public Administration Review*, 50(1), 35-41.
- Dolowitz, D. P., & Marsh, D. (2000). Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance* 13(1), 5-23.
- Forss, K., & Carlsson, J. (1997). The quest for quality—or can evaluation findings be trusted? *Evaluation*, 3(4), 481-501.

- Friedberg, A., & Lutrin, C. (2004). Are the watchdogs really watching? Assessment of the audit institutions in the fifty United States of America. *Iyunim, Studies in State Audit*, 60, 91-119.
- Funkhouser, M. (2005, October). *Accountability, performance and performance auditing: Reconciling the views of scholars and auditors*. Paper presented at Accountable Governance, An International Research Colloquium, Belfast.
- General Accounting Office. (2003). *Government Auditing Standards* (GAO Publication No. GAO-03-673G). Washington, DC: U.S Government Printing Office.
- Greathouse, F. L., & Funkhouser, M. (1988). Audit standards and performance auditing in state government. *The Government Accountants Journal*, 36(4): 56-60.
- Green, A. (1984). The role of evaluation in legislative decision making. *Public Administration Review*, 44(3), 265-267.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Gunvaldsen, J., & Karlsen, R. (1999). The auditor as an evaluator. *Evaluation*, 5(4): 458-467.
- Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, 97, 69-81.
- Hoogerwerf, A. (1992). Policy evaluation and government in the Netherlands: Meta evaluation research as one of the solutions. In J. Mayne, M. L. Bemelans-Videc, J. Hudson, & R. Conner (Eds.), *Advancing Public Policy Evaluation: Learning from International Experiences* (pp. 215-227). Amsterdam: Elsevier.
- Hoogerwerf, A., & Zoutendijk, D. C. (1990). *Naar een evaluatiesysteem voor overheidsbeleid* [Towards an evaluation system for public policies]. Enschede, Netherlands: University of Twente Press.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for Evaluations of Educational Programs, Projects, and Materials*. New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.
- Jonas, R. K. (1999). Against the whim: State legislatures' use of program evaluation. *New Directions for Evaluation*, 81, 3-10.
- Jones, R. (1987, July). Keeping an eye on state agencies. *State Legislatures*, 13, 20-23.

- Knighton, L. M. (1967). *The performance post audit in state government*. East Lansing, MI: Michigan State University, Bureau of Business and Economic Research.
- Kruse, S. (2005). Meta-evaluations of NGO experience: Results and challenges. In G. K. Pitman, O. N. Feinstein, & G. K. Ingram (Eds.), *Evaluating development effectiveness* (pp. 109-1127). New Brunswick, NJ: Transaction.
- Larson, R., & Berliner, L. (1983). On evaluating evaluations. *Policy Sciences*, 16(2), 147-163.
- Larson, R. C., Bier, V. M., Kaplan, E. H., Mattingly, C., Eckels, T. J., Reichman, N., et al. (1979). *Interim analysis of 200 evaluations on criminal justice*. Cambridge, MA: Massachusetts Institute of Technology, Operations Research Center.
- Leeuw, F. L. (1992). Performance auditing and policy evaluation: Discussing similarities and dissimilarities. *The Canadian Journal of Program Evaluation*, 7(1): 53-68.
- Leeuw, F. L. (1998). Doelmatigheidsonderzoek van de Rekenkamer als regelgeleide organisatiekunde met een rechtssociologisch tintje? *Recht der Werkelijkheid*, 14, 35-71.
- Leeuw, F. L., & Cooksy, L. J. (2005). Evaluating the performance of development agencies: The role of meta-evaluations. In G. K. Pitman, O. N. Feinstein, & G. K. Ingram (Eds.), *Evaluating development effectiveness* (pp. 95-108). New Brunswick, NJ: Transaction.
- Lipsey, M. W. (1988). *Codebook: Studies with delinquency outcome variables*. Unpublished Manuscript.
- Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. *New Directions for Program Evaluation*, 27.
- Louisiana Performance Audit Program Act. (1995). *Louisiana Revised Statutes* 24:522.
- McIntire, P. W., & Glaze, A. S. (1999). The use of systems modeling in legislative program evaluation. *New Directions for Evaluation*, 81, 45-60.
- Mohan, R. (1997, May). Legislative evaluators: A diverse groups of professionals. *NLPES News*. (Available at <http://www.ncsl.org/programs/nlpes/news/news0997.htm>)
- National Legislative Program Evaluation Society. (2004). *Ensuring the public trust: How program policy evaluation serves state legislatures*. Retrieved 1/30/06 from http://www.oppaga.state.fl.us/monitor/reports/pdf/ppe_publictrust_2004.pdf.

- National Legislative Performance Evaluation Society. (2007). NLPES member states and key contacts. Retrieved 10/30/07 from <http://www.ncsl.org/programs/nlpes/about/members/members.htm>
- Newcomer, K.E. (1994). Opportunities and incentives for improving program quality: Auditing and evaluating. *Public Administration Review*, 54(2), 147-154.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd. ed.). Thousand Oaks, CA: Sage.
- Pawson, R. (2002a). Evidence-based policy: In search of a method. *Evaluation*, 8(2), 157-181.
- Pawson, R. (2002b). Evidence-based policy: The promise of 'realist synthesis'. *Evaluation*, 8(3), 340-358.
- Pollitt, C. & Summa, H. (1996). Performance audit and evaluation: Similar tools, different relationships? *New Directions for Program Evaluation*, 71, 29-50.
- Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research*. Thousand Oaks, CA: Sage.
- Reichardt, C. S., & Rallis, S. F. (Eds.). (1994). The qualitative-quantitative debate: New perspectives. *New Directions for Program Evaluation*, 61.
- Rist, R. C. (1989). Management accountability: The signals sent by auditing and evaluation. *Journal of Public Policy*, 9(3): 355-369.
- Roberts, S., & Pollitt, C. (1994). Audit or evaluation? A national audit office VFM study. *Public Administration*, 72(Winter): 527-549.
- Sanders, J. (1999). Metaevaluation of "The effectiveness of comprehensive, case management interventions: Evidence from the national evaluation of the comprehensive child development program". *American Journal of Evaluation*, 20(3), 577-582.
- Schwartz, R. (1999). Coping with the effectiveness dilemma: Strategies adopted by state auditors. *International Review of Administrative Sciences*, 65: 511-526.
- Schwartz, R., & Mayne, J. (2005). Assuring the quality of evaluative information: Theory and practice. *Evaluation and Program Planning*, 28, 1-14.

- Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23(4), 387-419.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, 2, 36-38.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Scriven, M. (2006). *Key Evaluation Checklist*. Retrieved 7/14/07 from http://www.wmich.edu/evalctr/checklists/kec_june06.pdf.
- Sedgwick, S. (1993). When does an audit become an evaluation? *Canberra Bulletin of Public Administration*, 73, 148-151.
- Shadish, W. R. (Ed.). (1995). The quantitative-qualitative debates: DeKuhnifying the conceptual context. *Evaluation and Program Planning*, 18, 47-49.
- Song, K. (2003, May). *Towards Program Evaluation beyond Performance Audit: The Korean Experience*. Paper presented at the 11th International Anti-Corruption Conference, Seoul.
- Studer, S., Spitz, B., & Burt, M. R. (1981). Optimizing legislative program evaluation. In D. J. Palumbo, S. B. Fawcett, & P. Wright (Eds.), *Evaluating and optimizing public policy* (pp. 165-175). Lexington, MA: Lexington Books.
- Stufflebeam, D. L. (2000a). The CIPP model for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models* (2nd ed., pp. 279-318). Boston: Kluwer Academic Publishers.
- Stufflebeam, D. L. (2000b). The methodology of metaevaluation as reflected in metaevaluations by the Western Michigan University Evaluation Center. *Journal of Personnel Evaluation in Education*, 14(1) 95-125.
- Stufflebeam, D. L. (2001a). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D. L. (2001b). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22(1), 71-79.
- Stutzman, M. P., & Rainey, H. G. (1981). Program evaluation: State administrators' perspectives. In L. Datta (Ed.), *Evaluation in change: Meeting new government needs* (pp. 81-88). Beverly Hills, CA: Sage.

- Thomas, R., Gooking, K., Keating, B., & Whitener, V. (1999). Lessons learned from evaluating the feasibility of privatizing government services. *New Directions for Evaluation*, 81, 23-31.
- Van de Vall, M. (1980). *Sociaal beleidsonderzoek: een professioneel paradigma* [Social policy research: A professional paradigm]. Alphen aan den Rijn: Samson.
- VanLandingham, G. R. (2006). *A voice crying in the wilderness: Legislative oversight agencies efforts to achieve utilization*. Unpublished doctoral dissertation, Florida State University, Tallahassee.
- VanLandingham, G., Kinton, C., Boerner, B., & Zelio, J. (2000). Ensuring the public trust: How program policy evaluation is serving state legislatures. Retrieved 1/30/06 from <http://www.oppaga.state.fl.us/monitor/ppesurvey.html>.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick, NJ: Transaction.
- Walker, W. E. (1985). The impact of General Accounting Office program evaluations on government. *Evaluation and Program Planning*, 359-366.
- Walton, K. S., & Brown, R. E. (1990). State legislators and state auditors: Is there an inherent role conflict? *Public Budgeting and Finance*, 10(1), 3-12.
- Wheat, E. M. (1991). The activist auditor: A new player in state and local politics. *Public Administration Review*, 51(5), 385-392.
- Yarbrough, D. B., Shulha, L. M., & Caruthers, F. (2004). Background and history of the joint committee's program evaluation standards. *New Directions for Evaluation*, 104, 15-30.

Appendix A

Metaevaluation Criteria

APPENDIX A

Metaevaluation Criteria

| Preliminaries |
|--|
| Statement of Scope and Objectives |
| Excellent: The report lists the specific agency/department and/or activities to be evaluated/audited. The report lists the specific objectives of the audit/evaluation and how the objectives were determined (by legislative direction, statute, etc.). |
| Good: The report lists the specific agency/department and/or activities to be evaluated/audited. The report lists the specific objectives of the audit/evaluation but not how the objectives were determined. |
| Fair: The report specifies the agency/ department/activity evaluated or the specific objectives but not both or lists nonspecific objectives (such as “operating in the public interest”). |
| Poor: The report fails to specify both the agency/department/activities evaluated/audited and the objectives. |
| Not Addressed: Not applicable to this criterion. |
| Identification of Values |
| Excellent: The report clearly identifies the values used to determine the evaluative conclusions and how these values were chosen (the source of these values). |
| Good: The report clearly identifies the values used to determine the evaluative conclusions but not how these values were chosen (the source of these values). |
| Fair: The report implies the values used without clearly identifying them. |
| Poor: The report fails to address the values used. |
| Not Addressed: Not applicable to this criterion. |
| Quality Control |
| Excellent: The report clearly identifies sufficient internal quality control measures and external peer reviews conducted of its organization. |
| Good: The report clearly identifies sufficient internal quality control measures but fails to identify external peer reviews conducted of its organization. |
| Fair: The report identifies limited internal quality control measures and no external organizational reviews. |
| Poor: The report identifies no internal quality control measures and no external organizational reviews. |
| Not Addressed: Not applicable to this criterion. |

APPENDIX A—continued

Statement of Standards Compliance

Excellent: The report prominently states what standards (PES, GAGAS), if any, guide its conduct (for example, if there is a methodology section, the standards followed are listed there as well as in the introduction, summary, or letter of transmittal).

Good: The report states what standards, if any, guide its conduct but do not identify these standards in the methodology section.

Fair: The report refers to adherence to general professional standards or “government auditing standards” but does not specify which standards.

Poor: The report does not address standards.

Not Addressed: Not applicable to this criterion.

Stakeholder Identification

Excellent: The report clearly identifies stakeholders for the program/agency/policy under review, including all of the following: consumers of the evaluand, staff working on the evaluand, decision-makers about the evaluand and funders of the evaluand.

Good: The report clearly identifies stakeholders for the program/agency/policy under review, including more than one but not all of the following: consumers of the evaluand, staff working on the evaluand, decision-makers about the evaluand and funders of the evaluand.

Fair: The report identifies limited stakeholders or identifies only an overly broad group of stakeholders (e.g. the people to the state).

Poor: The report fails to identify any stakeholders.

Not Addressed: Not applicable to this criterion.

Appendix A—continued

| Methodology |
|--|
| Complete |
| Excellent: The report covers all of the areas it claims to cover in the scope and objectives sections. |
| Good: The report covers all but a single small area mentioned in the scope and objectives section. |
| Fair: The report covers most areas it claims to cover in the scope and objectives section. |
| Poor: The report fails to cover most of the areas it claims to cover in the scope and objectives section. |
| Not Addressed: Not applicable to this criterion. |
| Accurate |
| Excellent: Agency comments are included and the agency did not dispute the accuracy of the information reported. |
| Good: The agency disputed the accuracy of some information but the report convincingly rebuts this. |
| Fair: The agency convincingly disputes the accuracy of some minor information in the report. |
| Poor: Agency comments convincingly dispute report's accuracy. |
| Not Addressed: The type of report or subject matter does not lend itself to the inclusion of agency comments. |
| Defensible |
| Excellent: There was "sufficient, competent and relevant evidence . . . obtained to provide a reasonable basis for the auditors' findings and conclusions." |
| Good: There was "sufficient, competent and relevant evidence . . . obtained to provide a reasonable basis for the auditors' findings" but the overall conclusions go beyond the individual findings. |
| Fair: There was competent and relevant information presented but the conclusions go beyond what that evidence supports. |
| Poor: There was not sufficient, competent and relevant information presented to provide a reasonable basis for the findings. |
| Not Addressed: Not applicable to this criterion. |

Appendix A—continued

Human Subjects' Rights

Excellent: There is no evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.) and the report identifies procedures designed to ensure the rights of human subjects.

Good: There is no evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.).

Fair: There is some evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.).

Poor: There is strong evidence that the report authors violated subjects rights (disclosed confidentiality, presented negative information without opportunity for rebuttal, etc.).

Not Addressed: The report did not include activities that could undermine the rights of human subjects (review of previous literature, etc.)

Comparisons

Excellent: The report competently addresses evaluations/audits of similar programs/agencies.

Good: The report refers to evaluations/audits of similar programs/agencies but doesn't address them fully.

Fair: The report addresses similar programs/agencies but doesn't address evaluations/audits of these programs.

Poor: The report fails to refer to evaluations/audits of similar programs/agencies.

Not Addressed: Not applicable to this criterion.

Appendix A—continued

| Results |
|--|
| Context Analysis |
| Excellent: The report presents the context in which the program/policy/agency exists “in enough detail, so that its likely influences on the program can be identified.” |
| Good: The report presents the context in which the program/policy/agency exists but does not address the context’s influences. |
| Fair: The context is presented in insufficient detail. |
| Poor: The report doesn’t present, or only marginally presents, the context. |
| Not Addressed: Not applicable to this criterion. |
| Clear |
| Excellent: The report is (1) easy to read and understand, (2) technical terms, abbreviations and acronyms are explained, and (3) the material is logically organized. |
| Good: The report is easy to read and understand and logically organized but technical terms, abbreviations and acronyms are not explained. |
| Fair: The report is easy to read and understand but is not logically organized and technical terms, abbreviations and acronyms are not explained. |
| Poor: The report is not easy to read and understand. |
| Not Addressed: Not applicable to this criterion. |
| Concise |
| Excellent: The report is only as long as is necessary to convey and support conclusions. Repetition and extraneous detail are avoided. |
| Good: The report is only as long as is necessary to convey and support conclusions but includes some repetition and extraneous detail. |
| Fair: The report is longer than necessary to convey and support conclusions. |
| Poor: The report includes extensive repetition and/or extraneous detail. |
| Not Addressed: Not applicable to this criterion. |

Appendix A—continued

Findings and Conclusions

Excellent: Findings are supported by credible evidence related to the report's objectives. Conclusions are reported as "logical inferences from the findings" not simply a restatement of findings.

Good: Findings are supported, but some conclusions are simply a restatement of findings.

Fair: Findings are supported, but conclusions are mostly a restatement of findings.

Poor: Findings are insufficiently supported and conclusions are mostly a restatement of findings.

Not Addressed: Not applicable to this criterion.

Recommendations

Excellent: All recommendations presented are specific, measurable, and logically follow findings and conclusions.

Good: Most recommendations presented are specific, measurable, and logically follow findings and conclusions.

Fair: Recommendations logically follow findings and conclusions but are unspecific or not measurable.

Poor: Recommendations do not logically follow findings and conclusions.

Not Addressed: Not applicable to this criterion.

Timely

Excellent: The report was completed within the requirement (if any) of the authorization.

Good: The report's time requirement was extended by the report's authorizer and the report was completed within this extended deadline.

Fair: The report was not completed within the requirements but legitimate reasons are given for this.

Poor: The report was not completed with the requirements with no reasonable explanation.

Not Addressed: Not applicable to this criterion.

Appendix B

List of LPE Sample Reports

Appendix B

LIST OF LPE SAMPLE REPORTS

| State | Name of Office | Report Title | Year | Type |
|-------------|---|---|------|--------|
| Alaska | Division of Legislative Audit | Department of Commerce, Community and Economic Development, Board of Veterinary Examiners, Sunset Review | 2005 | Sunset |
| Alaska | Division of Legislative Audit | Department of Public Safety, Division of Fish and Wildlife Protection, Aircraft Section | 2002 | Other |
| Alaska | Division of Legislative Audit | Department of Revenue, Division of Treasury, Treasury Revenue Management System | 2002 | Other |
| Alaska | Division of Legislative Audit | Department of Transportation and Public Facilities, Southeast Region Construction, Maintenance & Operations, Select Administrative Issues | 2002 | Other |
| Arizona | Office of the Auditor General | Yuma Elementary School District | 2005 | PA |
| Arizona | Office of the Auditor General | Arizona Department of Commerce | 2003 | Sunset |
| Arizona | Office of the Auditor General | Arizona Department of Education—Early Childhood Block Grant Program | 2001 | Eval |
| Arizona | Office of the Auditor General | Arizona Game and Fish Commission—Heritage Fund | 2001 | PA |
| Arkansas | Division of Legislative Audit | Department of Finance & Administration Office of Child Support Enforcement Performance Audit | 2004 | PA |
| California | Legislative Analyst's Office | An Overview of California's Enterprise Zone Hiring Credit | 2003 | Eval |
| California | Legislative Analyst's Office | Water Special Districts: A Look at Governance and Public Participation | 2002 | Eval |
| Colorado | Office of the State Auditor | Public Employees' Retirement Association of Colorado (PERA) Administrative Expenses, Performance Audit | 2005 | PA |
| Colorado | Office of the State Auditor | Division of Central Services, Performance Audit, Department of Personnel & Administration | 2003 | PA |
| Colorado | Office of the State Auditor | Motor Carrier Regulation, Public Utilities Commission and Department of Revenue, Performance Audit, Department of Regulatory Agencies | 2003 | PA |
| Connecticut | Program Review and Investigations Committee | Energy Availability in Connecticut | 2001 | Other |

Appendix B—continued

| State | Name of Office | Report Title | Year | Type |
|---------|--|---|------|------|
| Florida | Office of Program Policy Analysis and Government Accountability (OPPAGA) | A Medicaid Buy-in Program Would Increase Health Care Access for the Uninsured But Also Increase State Costs | 2005 | Eval |
| Florida | OPPAGA | OPPAGA Report: Most Adult Protection Working Agreements Signed, But Have Had Limited Effectiveness | 2004 | Eval |
| Florida | OPPAGA | Justification Review: Environmental Laboratory Privatization Feasible; Cost Savings Are Uncertain | 2001 | Eval |
| Florida | OPPAGA | Justification Review: Financial Accountability for Public Funds Program Is Performing Well | 2001 | Eval |
| Georgia | Department of Audits and Accounts | DJJ Health Care Services | 2005 | Eval |
| Georgia | Department of Audits and Accounts | Crime Victim Services Part II: Compensation and Assistance Services | 2002 | Eval |
| Georgia | Department of Audits and Accounts | Property Forfeitures Under State and Federal Drug Laws | 2002 | Eval |
| Hawaii | Office of the Auditor | Audit of the Deposit Beverage Container Program | 2005 | PA |
| Hawaii | Office of the Auditor | Audit of Selected Procurement, Human Resource, and Fiscal Issues of the Hawaii Health Systems Corporation | 2004 | PA |
| Hawaii | Office of the Auditor | Review of Revolving Funds, Trust Funds, & Trust Accounts of the Office of the Governor, Office of Lieutenant Governor, Department of Education & Hawaii State Public Library System, & Office of Hawaiian Affairs | 2002 | PA |
| Hawaii | Office of the Auditor | A Review and Assessment of the Department of Education's Development of Educational Standards | 2001 | PA |
| Idaho | Office of Performance Evaluations | Idaho Department of Fish and Game: Opportunities Exist to Improve Lands Program and Strengthen Public Participation Efforts | 2001 | Eval |

Appendix B—continued

| State | Name of Office | Report Title | Year | Type |
|-----------|------------------------------------|--|------|-------|
| Illinois | Auditor General | Illinois State Board of Education and Other State Agencies Providing Funding to Illinois' Regional Offices of Education | 2001 | PA |
| Indiana | Legislative Services Agency | Issues Relating to Recycling and Solid Waste Management Programs | 2003 | Eval |
| Kansas | Legislative Division of Post Audit | Highway Construction: Reviewing KDOT's Plans for Construction on Highway 183 South of Plainville | 2005 | PA |
| Kansas | Legislative Division of Post Audit | Register of Deeds Technology Fund: Reviewing the Amounts Collected and the Uses of Those Moneys | 2004 | PA |
| Kansas | Legislative Division of Post Audit | Financing Local Governments: Determining How to Avoid Future Problems Caused by State Revenue Shortfalls | 2003 | PA |
| Kansas | Legislative Division of Post Audit | Highway Construction Change Orders: Reviewing Costs Associated with Construction on Highway 36 Near Marysville | 2003 | PA |
| Kansas | Legislative Division of Post Audit | Local Government Reorganization: Assessing the Potential for Improving Cooperation and Reducing Duplication | 2003 | PA |
| Kansas | Legislative Division of Post Audit | Department of Agriculture: Reviewing the Water Structures Program | 2002 | PA |
| Kansas | Legislative Division of Post Audit | How Including Previously Unreported Moneys Spent on K-12 Education Affects Kansas' Rankings in Nationally Published Statistics | 2002 | PA |
| Kansas | Legislative Division of Post Audit | Department of Agriculture: A K-GOAL Audit of the Kansas Pesticide and Fertilizer Program | 2001 | PA |
| Kentucky | Legislative Research Commission | Uncollected Revenues and Improper Payments Cost Kentucky Millions of Dollars a Year | 2004 | Other |
| Louisiana | Legislative Auditor | Department of Transportation and Development, Construction Division | 2005 | PA |
| Louisiana | Legislative Auditor | Executive Department Grant Programs | 2002 | PA |
| Maryland | Department of Legislative Services | Department of Budget and Management - Managing for Results Initiative | 2004 | PA |

Appendix B—continued

| State | Name of Office | Report Title | Year | Type |
|---------------|---|---|------|------|
| Massachusetts | Legislative Post Audit and Oversight Bureau | Losing Credibility | 2002 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Michigan State University | 2005 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Office of Community Corrections | 2005 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Selected Educational Databases Maintained by the Center for Educational Performance and Information | 2005 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Clean Michigan Initiative, Natural Resources Protection Programs | 2004 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Human Resources Management Network (HRMN) Self-Service | 2004 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Management Information Database | 2004 | PA |
| Michigan | Office of the Auditor General | Performance and Financial Audit - Bureau of Workers' Disability Compensation | 2002 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Division of HIV/AIDS - STD | 2002 | PA |
| Michigan | Office of the Auditor General | Performance Audit - Fire Marshal Division | 2002 | PA |
| Michigan | Office of the Auditor General | Performance Audit - West Shore Community College | 2001 | PA |
| Minnesota | Office of the Legislative Auditor | Energy Conservation Improvement Program | 2005 | Eval |
| Minnesota | Office of the Legislative Auditor | Preserving Housing: A Best Practices Review | 2003 | Eval |
| Mississippi | Joint Committee on Performance Evaluation and Expenditure Review (PEER) | A Review of the Legality of the Chancery Court Order Directing Annual Payments of Twenty Million Dollars in Perpetuity to the Partnership for a Healthy Mississippi | 2003 | Eval |
| Mississippi | PEER Committee | A Review of the Mississippi Adequate Education Program Funding Process | 2002 | Eval |
| Missouri | Joint Committee on Legislative Research | Program Evaluation: Professional Athletes and Entertainers Tax | 2002 | Eval |
| Montana | Legislative Audit Division | Highway and Bridge Construction Contract Claim Process | 2003 | PA |

Appendix B—continued

| State | Name of Office | Report Title | Year | Type |
|----------------|--|---|------|-------|
| Montana | Legislative Audit Division | Petroleum Tank Release Compensation Fund | 2003 | PA |
| Nebraska | Legislative Research Division | Nebraska Habitat Fund | 2002 | Eval |
| Nevada | Legislative Council Bureau | Department of Transportation, Highway Planning and Real Property Management | 2002 | PA |
| New Hampshire | Office of the Legislative Budget Assistant | Department of Education, Adequate Education Grant Data | 2004 | PA |
| New Mexico | Legislative Finance Committee | PBB Evaluation of Office of Cultural Affairs | 2002 | Eval |
| New Mexico | Legislative Finance Committee | State Engineer - Office of State Engineer and Interstate Stream Commission Audit Operations | 2002 | PA |
| New York | Committee on Oversight, Analysis and Investigation | No Room In The Playground: A Report Examining Playground Space in New York City Elementary Schools | 2003 | Other |
| Ohio | Legislative Office of Education Oversight | The Operating Costs of Ohio's eCommunity Schools | 2005 | Eval |
| Oregon | Legislative Fiscal Office | Audit of Oregon's Programs for Children and Families: Evaluation of Local Coordinated Comprehensive Planning | 2002 | Other |
| Pennsylvania | Legislative Budget and Finance Committee | The Status of Project Link-to-Learn and Other Educational Technology Initiatives | 2003 | Other |
| Rhode Island | Auditor General | Connect CARRE Performance Audit | 2003 | PA |
| South Carolina | Legislative Audit Council | A Review of Agency Restructuring and the Business Enterprise Program at the South Carolina Commission for the Blind | 2002 | PA |
| South Carolina | Legislative Audit Council | A Review of Higher Education Performance Funding | 2001 | PA |
| Tennessee | Comptroller of the Treasury, Division of State Audit | University of Tennessee Board of Trustees | 2005 | PA |
| Tennessee | Comptroller of the Treasury, Division of State Audit | Labor and Workforce Development, Department of: Department of Labor and Workforce Development and Related Entities | 2004 | PA |

Appendix B—continued

| State | Name of Office | Report Title | Year | Type |
|---------------|---|--|------|--------|
| Tennessee | Comptroller of the Treasury, Division of State Audit | Southwest Community Services Agency | 2004 | PA |
| Tennessee | Comptroller of the Treasury, Division of State Audit | Water Quality | 2001 | PA |
| Texas | Sunset Advisory Commission | Barber Examiners, Texas State Board of Pharmacy, Texas State Board of | 2004 | Sunset |
| Texas | Sunset Advisory Commission | An Audit Report on the Texas Education Agency's Administration of the Foundation School Program | 2004 | Sunset |
| Texas | State Auditor | Workforce Commission, Texas | 2003 | PA |
| Texas | Sunset Advisory Commission | An Audit Report on Correctional Officer Staffing at the Department of Criminal Justice | 2002 | Sunset |
| Texas | State Auditor | | 2001 | PA |
| Utah | Legislative Auditor General Joint Legislative Audit Review Commission | A Review of Decision-making Process Used to Obligate Federal TANF Surplus Funds | 2004 | Other |
| Virginia | Joint Legislative Audit Review Committee (JLARC) | Special Report: The Secretarial System | 2002 | Other |
| Washington | JLARC | HVAC/R Licensing & Testing Requirements Study | 2005 | Other |
| Washington | Legislative Performance Evaluation and Research Division (PERD) | Overview of Environmental Permitting for Transportation Projects | 2005 | Other |
| West Virginia | PERD | Prosecuting Attorneys Institute | 2005 | Other |
| West Virginia | PERD | West Virginia Conservation Agency | 2005 | Other |
| West Virginia | PERD | Board of Risk and Insurance Management | 2004 | Other |
| West Virginia | PERD | PEIA Finance Board | 2004 | Other |
| West Virginia | PERD | Potomac River Basin | 2003 | Other |
| West Virginia | PERD | Manufactured Housing Construction and Safety Board | 2002 | Other |
| West Virginia | PERD | Board of Acupuncture | 2001 | Other |
| West Virginia | PERD | Educational Broadcasting Authority | 2001 | Other |

Appendix B—continued

| State | Name of Office | Report Title | Year | Type |
|-----------|----------------------------|--|------|-------|
| Wisconsin | Legislative Audit Bureau | SeniorCare Eligibility | 2005 | Eval |
| Wisconsin | Legislative Audit Bureau | State of Wisconsin Investment Board | 2004 | Eval |
| Wisconsin | Legislative Audit Bureau | Office of Justice Assistance | 2003 | Other |
| Wisconsin | Legislative Audit Bureau | Pesticide and Fertilizer Fees | 2003 | Other |
| Wisconsin | Legislative Audit Bureau | Professional Football Stadium District | 2002 | Other |
| Wisconsin | Legislative Audit Bureau | Prior Authorization for Therapy and Other Services | 2001 | Eval |
| Wisconsin | Legislative Audit Bureau | Prison Health Care | 2001 | Eval |
| Wyoming | Legislative Service Office | Wyoming Aeronautics Commission | 2002 | Eval |

Note. Eval = evaluation; PA = performance audit.

APPENDIX C

List of Non-LPE Sample Reports

Appendix C

List of Non-LPE Sample Reports

| State | Name of Office | Report Title | Year | Type |
|----------------|-----------------------------|---|------|-------|
| North Carolina | Office of the State Auditor | Drivers Licensing Process | 2005 | PA |
| North Carolina | Office of the State Auditor | Community Alternatives Program for Disabled Adults | 2004 | PA |
| North Carolina | Office of the State Auditor | State Energy Policy Council and the State Energy Office | 2004 | PA |
| North Carolina | Office of the State Auditor | State Aircraft Operations | 2003 | PA |
| North Dakota | State Auditor | North Dakota Veterans' Home | 2002 | PA |
| North Dakota | State Auditor | Service Payments for the Elderly and Disabled (SPED) | 2001 | PA |
| Oklahoma | State Auditor and Inspector | Department of Public Safety Performance Audit | 2004 | PA |
| Vermont | State Auditor | Elections: Development and Implementation of Statewide Voter Checklist System Could Be Improved | 2005 | PA |
| Vermont | State Auditor | Getting There from Here: Recommendations to Modernize the Permit Process and Improve Financial Controls at the Department of Environmental Conservation | 2002 | Other |
| Vermont | State Auditor | Vermont's Restitution System: Failing to Pay the Victim | 2001 | Other |

Note. PA = performance audit.

APPENDIX D

List of Louisiana LPE Sample Reports

APPENDIX D

LIST OF LOUISIANA LPE SAMPLE REPORTS

| State | Name of Office | Report Title | Year | Type |
|-----------|---------------------|---|------|------|
| Louisiana | Legislative Auditor | Louisiana Housing Finance Authority, Administration of Energy Assistance Programs | 2005 | PA |
| Louisiana | Legislative Auditor | Department of Economic Development, Financial Assistance Programs | 2004 | PA |
| Louisiana | Legislative Auditor | Quality Jobs Program | 2004 | PA |
| Louisiana | Legislative Auditor | Louisiana Community and Technical College System | 2003 | PA |
| Louisiana | Legislative Auditor | Department of Economic Development - Business Tax Incentives | 2002 | PA |
| Louisiana | Legislative Auditor | Governor's Office of Women's Services | 2002 | PA |
| Louisiana | Legislative Auditor | Louisiana Tax Commission Ratio Studies | 2002 | PA |
| Louisiana | Legislative Auditor | Department of Economic Development, Office of Financial Institutions, A Review of the Louisiana CAPCO Program | 2001 | PA |

Note. PA = performance audit.