

An Evaluation Use Framework and Empirical Assessment

Laura R. Peck
Lindsey M. Gorzalski
*Arizona State University
School of Public Affairs*

Background: Research on evaluation use focuses on putting evaluation recommendations into practice. Prior theoretical research proposes varied frameworks for understanding the use (or lack) of program evaluation results.

Purpose: Our purpose is to create and test a single, integrated framework for understanding evaluation use. This article relies on prior theoretical research regarding categories of utilization, typologies of recommendations, and factors affecting utilization to frame an empirical study of evaluation use that then tests the integrated theory.

Setting: The empirical part of the article draws on post-evaluation interviews with sixteen agencies that have engaged in evaluation research.

Subjects: The agencies are mostly local non-profits, but the sample also includes a state agency, a city agency, and two university-community partnerships. All agencies had undergone a program evaluation between 2003 and 2006.

Intervention: Having participated in an evaluation is the main “intervention” of interest in this article, in which we consider the relationship between evaluation use theory and empirical evidence on the topic.

Research Design: A qualitative approach, our research design involved examining each of the sixteen agencies within two years of their having been evaluated.

Data Collection and Analysis: Data collection included structured in-person interviews with at least one key

informant in each agency. In addition, a short, closed-ended survey was administered to research participants. Interview data were analyzed using content analysis of themes and grouping agencies according to their evaluation’s outcomes (favorable or not) and experiences. Survey data were analyzed with simple descriptive statistics and similarly involved a subgroup analysis, according to agencies’ reported use of evaluation.

Findings: Most evaluation use is conceptual, and few agencies studied actually implemented specific evaluation recommendations. Agencies perceived recommendations as changes to rules and structure, which theory and prior research suggest suppresses use. An important human factor that influenced evaluation use was minimal post-evaluation interaction with evaluators. Some long-term influence of the evaluation is evident, but only as conceptual and not as instrumental. In fact, very little instrumental use existed in this empirical assessment.

Conclusions: Evidence suggests that evaluation use in practice aligns with theory, specifically emphasizing the conceptual use dimension. The proposed integrated model of prior evaluation use theories may hold value for future theoretical and empirical work.

Keywords: program evaluation; evaluation use, utilization, influence; teaching, curriculum and instruction; organizational performance

Several volumes of *New Directions for Evaluation* have focused on evaluation use and misuse (Caracelli & Preskill, 2000; McLaughlin, Weber, Covert & Ingle, 1988; Stevens & Dial, 1994). In this research, the terms utilization, use, and influence are often used, sometimes interchangeably, to suggest the ways in which evaluations—both specifically and generally—affect organizational operations, decisions and outcomes (Kirkhart, 2000; Patton, 1985; Weiss, 1981). Inherently, program evaluation and policy analysis are applied fields that are motivated by the possibility that some person, agency, or community will make use of the information. Although perhaps academically interesting, the results of policy and program evaluations are most relevant in the applied world. In other words, *the reason* to conduct an evaluation is, in the end, to help improve program operations and/or outcomes, generate knowledge, and/or to inform (presumably better) decision making (Preskill & Caracelli, 1997). Using the results of evaluations allows us to garner support for effective policies and programs, redirect scarce resources away from ineffective ones, and improve the effectiveness of programs in response to social need. Whether Program A has a 16 percent impact on outcome X is of little interest unless that information is used to make better decisions about the program, its replication, or the other programs in its general policy arena.

A trend noted in the utilization literature is the lack of use of program evaluation results. With the resources and time allocated to program evaluation, it is critical to understand how evaluations' processes, results and recommendations can be better utilized and implemented.¹ Without such knowledge of how to make program evaluation useful, the field, though important and necessary for development and success, may not continue to be valued in the public and nonprofit sector.

¹ Despite the continuing debate about identifying the proper term for evaluation use, we choose to continue the interchangeable use of the terms "use" and "utilization."

The goal of this research project is to combine existing frameworks of evaluation utilization—namely those developed by Alkin (1985), Downs (1967), Kirkhart (2000), and Taut and Alkin (2003)—and uncover possible relationships among framework factors through an empirical assessment of evaluation use. These theories each describe a framework of evaluation use, yet no single framework captures all the necessary dimensions, nor are the proposed frameworks empirically tested (with the exception of Taut & Alkin, 2003). Data for our study come from sixteen program evaluations conducted in the context of a graduate course on program evaluation. In-person structured interviews and surveys were completed with each program evaluated to understand how the agency had used the evaluation after at least six months and up to two years. Consistent with previous research is our observation that most evaluation use is conceptual, happening during the process of the evaluation. Initial results demonstrate the difficulties agencies have in using evaluation results, leading us to surmise that it may be more feasible for evaluators to aim for conceptual use over the long-term. While recommendations to implement will continue to have value, these recommendations may be viewed as "ideas for change" rather than steadfast items to implement. We believe these and additional findings from our empirical assessment of evaluation use are important to the field.

The paper proceeds as follows: First we review existing conceptual frameworks of evaluation use and aim to integrate them into a single framework. We then pose our research questions and describe our data and methods. Using our integrated framework of evaluation use, we analyze our data and report results regarding the ways and extent to which our data fit the framework. In discussing and concluding, we consider implications for evaluation practice and programmatic use of evaluation.

Conceptual Framework

Prior theoretical research has examined evaluation use and has proposed varied frameworks for understanding the use (or lack thereof) of program evaluation results (Cousins & Leithwood, 1986; Kirkhart, 2000; Taut & Alkin, 2003). Generally, this body of research is interested in ways in which evaluators can increase the potential use of their work and the factors that affect the likelihood that evaluation might be used to affect program operations or policy decisions (see Shulha & Cousins, 1997) for a comprehensive review of the literature). In particular, we rely on prior research regarding categories of utilization, typologies of recommendations, and factors affecting utilization to frame our empirical study of evaluation use. These three frameworks are also compared along the dimension of time. This section describes each of these frameworks and expresses how our research will use them to explore data from our sample of evaluations.

Categories of Use

While scholars have debated the definition of evaluation use, some common categories of use exist: instrumental, conceptual, and persuasive or symbolic (Chelimsky, 1983; Leviton & Hughes, 1981; Smith, 1988). Instrumental use refers to the direct use of an evaluation's findings in decision making or problem solving. The use of recommendations is the largest indicator of an evaluation's achievement, and if recommendations are carried out by the evaluand, then the evaluation is considered successful.

In comparison, an evaluation's contribution to changes in thinking or to the general knowledge base is classified as conceptual use. Within conceptual use, the impact of the evaluation does not have to include the use of the formal recommendations. Conceptual use also encompasses the idea of process-based use, as Kirkhart (2000) notes. Process use is the

knowledge gained through the course of conducting a program evaluation. While this use may be viewed as being less important to evaluators (Preskill & Caracelli, 1997), its role in evaluation utilization has become a focus in current evaluation literature (Kirkhart, 2000; Shulha & Cousins, 1997).

Persuasive use, or symbolic use, refers to an evaluation's use to convince others of a political position. Prior theoretical work has asserted that instrumental use is rare and limited by a variety of contextual factors, whereas evaluation results tend to contribute conceptually to broader understanding in a field, over time influencing what we know and how we make decisions about policies and programs (Mowbray, 1988).

These categories of evaluation use mirror somewhat the results-based and process-based tension that exists regarding evaluation use. That is, the instrumental use of program evaluation tends to expect that the empirical results of an evaluation are what will be used directly, whereas the conceptual use of evaluation is more likely to reflect longer-term process-based influences.

Typologies of Recommendations

In influencing the level at which an evaluation is utilized, one must consider the type of change the evaluation recommends. Johnston (1988) applied a typology of organizational change from Downs (1967) to describe evaluation utilization. The four types (or categories) of change—behavioral, rules-based, structural, and purpose-based—theoretically all affect the likelihood of accepting recommendations. Behavior changes involve individuals' behaviors and are generally in sync with existing organizational values and practices. Changes to rules include practices regarding how individuals should act in accordance with the formal regulations set by the organization. Structural changes include changes to the hierarchies, power, and income of an organization. Purpose-based changes involve altering the overall

mission and underlying values of the organization. According to Johnston (1988), behavioral changes are more likely to be accepted, whereas recommendations that change the purpose of the organization are least likely to be accepted. Johnston applied these categories of change to evaluations conducted by the General Accounting Office (GAO, now the Government Accountability Office) and found that 81 percent of the recommendations made that fell into the “behavior” category were thought to have a high likelihood of implementation. On the other hand, he did not examine the agency’s actual implementation of GAO recommendations, making his conclusions about use only speculative (Johnston, 1988). Both rules-based and structural changes fall between behavioral and purpose-based changes in term of their likelihood of being accepted.

Evaluation results that support the already-existing behaviors and beliefs of an organization have been found to be more likely accepted than recommendations at odds with an organization’s culture (Cousins & Leithwood, 1986). Marsh and Glassick (1988) studied how various types of recommendations made in an evaluation led to whether or not a recommendation would be utilized. They noted how considering context made recommendations more readily accepted.

Factors Affecting Utilization

In the “factors affecting utilization” framework, evaluators can capitalize on specific external factors to create a better environment for the implementation of their evaluation work. Originally developed by Alkin (1985), these factors include a human element, the context of the evaluation, and the characteristics of the evaluation itself. The human factor measures the personal characteristics of the evaluator, such as communication skills and competence. Credibility of the evaluator and communication with stakeholders have been demonstrated to

affect evaluation use (Cousins & Leithwood, 1986; Dibella, 1990).

Evaluation-specific factors include how the evaluation was conducted and can range from the choice of data collection to the scientific validity of the evaluation report. Cousins and Leithwood (1986) note the importance of a sound methodology and useful data in utilization, while Oman and Chitwood (1984) demonstrate that qualitative, descriptive evaluations are used more than quantitative, statistic-heavy results. Both the role of the human factors and the evaluation characteristics contribute to the idea of evaluation as organizational change (Torres & Preskill, 2001). This concept of organizational change presents a new, more integrated approach to evaluation that promotes use and is included in our framework.

Contextual factors represent the environment in which the evaluation was conducted. Conditions such as the political climate or the administrative mood fall into the contextual factor category and can heavily influence how an evaluation is received (Chelimsky, 1986). The political context of an evaluation dictates whether or not recommendations will be used or any changes will be made to the existing program.

Taut and Alkin (2003) studied the barriers to implementation of external program evaluations by examining the effects of the three utilization-enhancing factors—human factors, context, and evaluation characteristics. In their study, Taut and Alkin utilized this framework to analyze empirically the barriers to implementing evaluation recommendations. They report that human factors pose barriers to implementation of recommendations, with issues of trust, fear, and interpersonal skills of the evaluators being frequently named as impediments.

Time Dimensions

In theorizing about evaluation use, time is divided into three categories: immediate, end-of-cycle, and long-term influence (Smith, 1988). Immediate influence happens concurrently with the evaluation process and can be viewed in terms of process-based and early results-based learning. End-of-cycle influence occurs at the culmination of the evaluation process. Long-term influence captures the learning and influence that evaluation process and results have over time. The results of a program evaluation can be timely to implement and absorb, thus making the idea of time critical to any complete framework of evaluation use.

Toward an Integrated Framework

This discussion makes clear that varied frameworks aim to explain the use of evaluation results. Cousins and Leithwood (1986) developed a framework based on evaluation quality, the credibility of the evaluators, relevance of information, quality of communication, the findings of the evaluation, and timeliness of the dissemination of the findings. Kirkhart (2000) presents an integrated theory of evaluation influence based on three dimensions of evaluation use, intention, source, and time. While each framework offers a unique perspective, we believe none is fully comprehensive. In addition, these frameworks have not have been adequately tested. The research conducted by Cousins and Leithwood (1986) used a meta-analysis of other evaluation studies, and Kirkhart's (2000) framework remains untested. Given that evaluators are concerned with creating a pre-evaluation plan that will maximize use (Preskill & Caracelli, 1997), a comprehensive, empirically-validated framework should be important to the field.

Our intent in combining existing frameworks of evaluation use is to uncover relationships between and among framework elements that can be tested through

examination of real-world evaluation practice and use. We hope to address the reasons why evaluations may not be utilized with the hope of directing the field in conducting useful evaluations. While existing research has addressed this issue in theory, our particular contribution is to test an integrated framework against data from program evaluations. Ginsburg and Rhett (2003) note a lack of research examples regarding how evaluations have actually been used. Our literature search revealed only two empirical studies (Cousins & Leithwood, 1986; Taut & Alkin, 2003) conducted in the past two decades. Our study responds to Ginsburg and Rhett's (2003) critique to begin to fill that knowledge gap.

Pulling together elements from others' frameworks of evaluation use, our proposed integrated framework, depicted in Figure 1, includes all of the elements others have hypothesized to matter. Specifically, we concur that Taut and Alkin's (2003) notions of human, contextual and evaluation factors matter to whether an evaluation's findings will be used. In addition, the type of change recommended (behavior, structure, rules, purpose) may determine the extent of the type of influence (instrumental, conceptual, and persuasive/symbolic). Drawing from Kirkhart (2000), we posit that the time horizon matters, too. What we have added to these disparate concepts is a conceptualization of how they overlap and interact. We expect that factors affecting utilization (i.e., human, evaluation, contextual) create a backdrop upon which the type of suggested change is influenced, thus leading to the kind of evaluation used, all which are impacted by the role of time. This model will thus predict likelihood of use. The type of change suggested is, to a certain extent, a subset of the evaluation factors; and type of use may interact with time to create a setting in which evaluation results are more or less likely to be used. The process is not linear; instead there are certain choices to be made at each stage of the evaluation process. What our framework

presents is a possible model of prediction; certain efforts made by evaluators in each of the areas will influence the extent to and ways in

which an evaluation is used. Figure 1 is our visualization of the integrated framework.

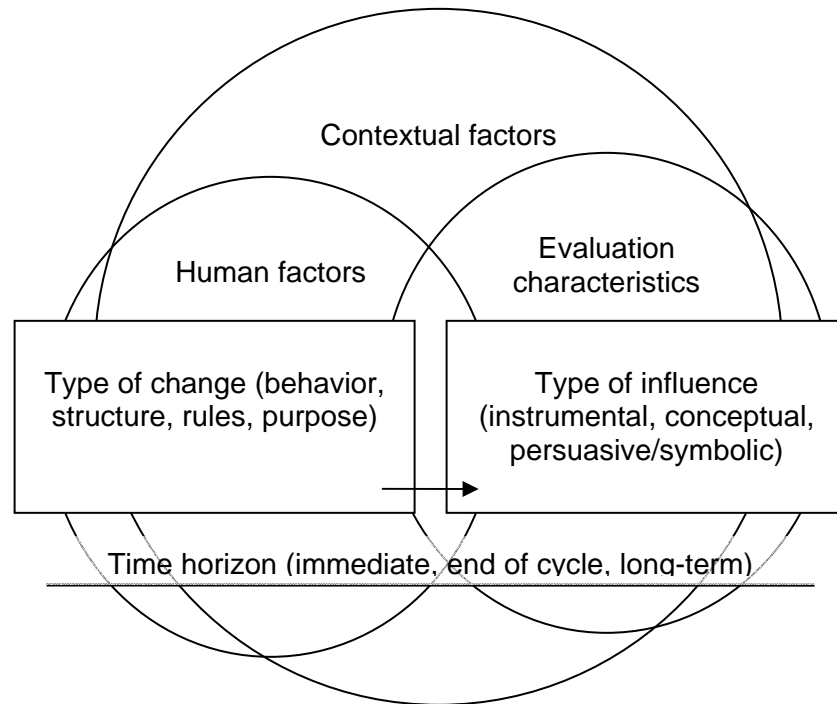


Figure 1. Integrated Framework for Evaluation Use

Research Questions

Reflecting on previous literature, our research aims to understand and illuminate how agencies use evaluation results in practice. Speculation regarding how to maximize evaluation use abounds, yet few empirical studies test those theories. Our initial descriptive research question is how do public and nonprofit agencies use the results of a program evaluation? Stemming from an answer to that question, we hope to learn to what extent the observed patterns of evaluation use align with theory about evaluation use.

Data and Methods

Data for this project come from sixteen program evaluations conducted in the context of a graduate course on program evaluation. The graduate course is offered annually and was first introduced in its current applied format in 2003. The sixteen evaluations considered here come from the 2003 through 2006 iterations of the course. The goal of the course is to produce both consumers and quality producers of program evaluation research through exposing students to real-world evaluation. Teams of students were assembled and given approximately three months to research their assigned agency, conduct the evaluation, and provide a comprehensive written report with recommendations. Many of the students in the

course had little or no previous experience with program evaluation, though the majority of the students were versed in research methods and statistical analysis. In brief, the course was structured to provide a primer on fundamental evaluation concepts and tools during three, eight-hour segments before students began to engage in the evaluation work. We assert that this level of knowledge was sufficient to classify the evaluation teams as solid beginners in their training. The professor's expert knowledge and experience in the field offered assurance to the partnering agencies that their evaluation questions would be competently researched. For more details on the course structure, content and goals, see Catlaw and Peck (2007).

A variety of agencies were recruited to participate in the course. Most of the partnering agencies were nonprofit organizations, but a few were university-community partnerships, and two were government agencies. In the years 2003, 2004, and 2005, professional contacts and word-of-mouth drove the partnering arrangements. In 2006, we extended a call for applications to local public and nonprofit

agencies to partner with the course in an evaluation. Although the resulting 2006 projects may have included clients with a better sense of what to expect from an evaluation, these projects and the partner agencies seem otherwise indistinguishable from the earlier projects. We have collected data from all but three of the projects over these four years. Data collection began in the fall of 2005 and extended over the following year. The projects not included in our sample are absent because of staff turnover (in two cases) and lack of time to participate (in one case). We have no reason to believe that the exclusion of these projects meaningfully biases our data.

For the purpose of maintaining some semblance of privacy regarding these organizations' work and experiences, we have created pseudonyms for the organizations and do not identify key informants, but the general nature of the organization or project work, as necessary for the analysis and interpretation, remains intact.

Table 1
Research Sample Description

Organization name, project	Type or organization, project	Evaluation question(s)	Type of evaluation and results
1. University Service Learning with Franklin School	University-community partnership	What effects does hand-on biology learning have on student outcomes?	Summative—favorable (some mixed)
2. City Historical Museum	City agency	How do public perception and organizational goals match?	Formative—favorable
3. State Department of Social Services	State agency	How do case management tools result in effective service targeting?	Summative—mixed
4. Community Food Bank	Nonprofit	What effect does food distribution to intermediary organizations have on hunger among target populations?	Formative—favorable Summative—favorable
5. Community Cultural Affairs, leadership training program	Nonprofit	What effect does leadership education have on civic engagement attitudes and practice?	Formative—unfavorable Summative—inconclusive

6.	Statewide School Partnership, targeted service provision	Nonprofit	What effect does providing low-income elementary schools with comprehensive health and human services have on student achievement?	Summative—mixed
7.	University Service Learning with Sunnyside School	University-community partnership	What are the knowledge and attitudinal effects of engaging in environmental service?	Formative—unfavorable Summative—inconclusive
8.	State Department of Child Care Services, statewide training program	State agency	What effects do contracted child care initiatives have on the quality or supply of child care?	Summative—very favorable
9.	Statewide School Partnership, Teachers for Tomorrow program	Nonprofit	What effect does teacher training have on teaching strategies used in the classroom?	Summative—favorable (some mixed)
10.	Statewide Community Action Group (CAG)	Nonprofit	To what extent have CAG's efforts impacted Food Stamp Program application and participation rates?	Summative—inconclusive
11.	Local Community Action Network (CAN)	Nonprofit	What are the outcomes of Local CAN's programs?	Formative—unfavorable
12.	Boys and Girls Agency, United Nations program	Nonprofit	What effects does the Model UN program have on its middle school student participants?	Summative—favorable (some mixed)
13.	10%, LGBTQ youth support program	Nonprofit	What impacts does 10% have on participating youth's well-being?	Formative—favorable Summative—favorable (though tentative)
14.	Youthplay, theater program	Nonprofit	What impacts does Youthplay's School Touring Program have on children and teachers?	Summative—favorable (though tentative)
15.	Feed Kids Afterschool - Community Food Bank Alliance	Nonprofit	To what extent is Feed Kids Afterschool meeting its objectives and having an effect on participants' academic achievement?	Formative—favorable Summative—inconclusive
16.	i-Mentor Online Youth Services	Nonprofit	How is i-Mentor's program implementation going, what preliminary outcomes has it achieved, and how might it measure impacts in the future?	Formative—favorable (some mixed) Summative—inconclusive

Note: Organizations are listed in chronological order, with numbers 1-3 having projects in the 2003 course, 4-7 having them in 2004, 8-12 having them in 2005, and 13-16 having them in 2006. Judgments about the evaluation results are the authors' assessments.

An in-person interview was completed with each agency contact person to collect information about how the agency used the evaluation. Questions were crafted to encourage response regarding all the different types of evaluation utilization and any possible barriers to utilization (see Appendix A for interview

guide). Although similar studies have used open-ended questions to illicit a more spontaneous response (Taut & Alkin, 2003), our study specifically looked to test certain assumptions, therefore structured questions were necessary. Each participant was provided, prior to the interview, with a short list of topics

and themes to be discussed during the interview in addition to an electronic copy of the original evaluation report. Because some projects had partnered with the course over two years prior to the interview, it was necessary to give each contact an opportunity to refresh his or her memory. The interviews were conducted by a nonbiased third party who did not participate in the original evaluation process.² Interviews were transcribed and coded thematically according to the conceptual framework categories. Each interview was examined with regard to type of change, type of influence, factors contributing to utilization, and time. Each time an interviewee mentioned a unique aspect of the framework, this was noted and combined with other like factors to create a frequency measure for each factor. The aggregate answers were analyzed using content analysis for various framework themes. We imported the interview transcript notes into a spreadsheet to facilitate sorting according to various dimensions of interest, including evaluation results (favorable or unfavorable), type of change, type of influence, utilization factors, and time as well as reported evaluation use.

In addition, a short closed-ended survey was administered (see Appendix B). The essential elements of the survey for this research include the following statements, to which respondents were asked the extent of their agreement or disagreement in a Likert response format:

- The recommendations given in the evaluation report were utilized.
- The findings of the evaluation report were favorable to my agency.
- The evaluation report was useful to my agency.

Data were coded such that “strongly disagree” earned -2 points, “disagree” earned -1,

being “neutral” earned 0, “agree” earned +1, and “strongly agree” earned +2. As a result, reported means that are positive in value are also favorable in meaning; whereas values that are negative are unfavorable in meaning, and a zero value reflects a neutral view. Simple descriptive statistics and crosstabulations were performed to assess the factors associated with evaluation use.

Analytic Findings

In this section, we first report on the summary findings from our closed-ended survey. We then report on our analysis of interview data, with an eye toward the ways in which our findings align (or do not align) with elements of the theoretical framework for evaluation use.

Survey Findings

Table 2 reports the summary results from the survey of project contacts. In instances where there was more than one informant, we averaged the responses for that project so that each project makes only one contribution to the analysis.

² In the case of one interview, the participant did previously know the interviewer. This participant was given the opportunity to have an alternate interviewer but chose not to.

Table 2
Summary of Survey Responses

Survey question: To what extent do you agree with...	Average	% Agree OR Strongly Agree	% Strongly Agree
Program evaluation is important to my agency.	1.84	100.0	81.3
Program evaluation can tell me whether or not a program is working as intended.	1.69	100.0	62.5
Participation of an agency or organization is critical in conducting a program evaluation.	1.97	100.0	93.8
The process of participating in a program evaluation affected my thoughts and feelings about my agency or program.	0.84	75.0	25.0
It was easy for me to work with the evaluation team.	1.23	86.7	33.3
I found the evaluation team to be knowledgeable about program evaluation.	1.20	93.3	20.0
I was confident in the competence of the evaluation team.	1.17	86.7	26.7
I was confident in the competence of the professor of the program evaluation course.	1.89	93.3	80.0
The findings of the evaluation report were favorable to my agency.	0.75	62.5	25.0
The recommendations given in the evaluation report were utilized.	0.80	75.0	12.5
I was satisfied with the report produced by the student evaluation team.	1.22	93.8	25.0
The evaluation report was useful to my agency.	1.41	93.8	43.8

Note: Values were coded as follows: that “strongly disagree” earned -2 points, “disagree” earned -1, being “neutral” earned 0, “agree” earned +1, and “strongly agree” earned +2. Reported averages that are positive in value are also favorable in meaning; whereas values that are negative are unfavorable in meaning.

In addition to these summary results, selected crosstabulations (see Table 3) reveal some tendencies associated with evaluation use. For example, two projects strongly agreed that the recommendations were utilized and another ten agreed that the recommendations were utilized (the remaining four were either neutral or disagreed), and there are some important differences between the utilizers and nonutilizers of the evaluation results. Specifically, those who used recommendations had found it easier to work with the evaluator and had more confidence in the team’s

knowledge and competence. The latter of these is one of the largest differences: While all of the utilizers had confidence in the team’s competence, just half of the nonutilizers had such confidence. Interestingly, those who used the recommendations agreed *less* that the findings were favorable to the agency. This may highlight that unfavorable evaluation results compel action in response. That said, those who used recommendations were more satisfied with the report and also agreed more that the evaluation report was useful.

Table 3
Comparison of Survey Responses between Utilizers and Nonutilizers

Survey question: To what extent do you agree with...	Average	% Agree OR Strongly Agree
<i>Among the twelve who agreed or strongly agreed that the recommendations given in the evaluation report were utilized...</i>		
It was easy for me to work with the evaluation team.	1.27	90.9
I found the evaluation team to be knowledgeable about program evaluation.	1.23	100.0
I was confident in the competence of the evaluation team.	1.27	100.0
I was confident in the competence of the professor of the evaluation course.	1.90	90.9
The findings of the evaluation report were favorable to my agency.	0.63	58.3
I was satisfied with the report produced by the student evaluation team.	1.29	100.0
The evaluation report was useful to my agency.	1.54	100.0
<i>Among the four who were neutral or disagreed that the recommendations given in the evaluation report were utilized...</i>		
It was easy for me to work with the evaluation team.	1.13	75.0
I found the evaluation team to be knowledgeable about program evaluation.	1.13	75.0
I was confident in the competence of the evaluation team.	0.88	50.0
I was confident in the competence of the professor of the evaluation course.	1.88	100.0
The findings of the evaluation report were favorable to my agency.	1.13	75.0
I was satisfied with the report produced by the student evaluation team.	1.00	75.0
The evaluation report was useful to my agency.	1.00	75.0
<i>Difference between USED and did NOT USE recommendations...</i>		
It was easy for me to work with the evaluation team.	0.15	15.9
I found the evaluation team to be knowledgeable about program evaluation.	0.10	25.0
I was confident in the competence of the evaluation team.	0.40	50.0
I was confident in the competence of the professor of the evaluation course.	0.02	-9.1
The findings of the evaluation report were favorable to my agency.	-0.50	-16.7
I was satisfied with the report produced by the student evaluation team.	0.29	25.0
The evaluation report was useful to my agency.	0.54	25.0

Note: Values were coded as follows: that “strongly disagree” earned -2 points, “disagree” earned -1, being “neutral” earned 0, “agree” earned +1, and “strongly agree” earned +2. Reported averages that are positive in value are also favorable in meaning; whereas values that are negative are unfavorable in meaning.

Interview Findings

Overall, we observed a great deal of variation in the responses of interview participants. Because each program evaluated was inherently different, such variance was expected. There are, though, several important themes for interpretation within the conceptual framework. In examining the role of use-influencing factors (human, evaluation characteristics, and context), the importance of human factors became evident. When asked about how confident they were with the evaluation team, the agency contacts varied on their answers. The positive responses included such statements as “very confident in team,” “group was a great team,” the team was energetic, competent, and ambitious,” and “the team asked a lot of

questions and spent a lot of time trying to understand the program.” More than half of the respondents voiced concerns about the competence of the evaluation team, providing statements like “[my] confidence level was mediocre at best,” “I was concerned with the members of evaluation team as [only] two members were actively involved,” “team did not have firm grasp of the program,” and “[it] seemed the team didn’t understand what needed to be done.” These statements support the findings noted in the survey that nonutilizers of evaluation recommendations were less confident with their evaluation team.

Despite the mixed results on the perceived competence of the evaluation teams, few agency contacts had complaints about communication. The positive responses consisted of statements

such as “team was professional and courteous,” “team was responsive, professional, and aware of customers needs,” and “I really felt I could communicate with the team leader.” Stakeholder involvement was also an area of satisfaction for most respondents, an interesting result given that most stakeholders and agency contacts had a small role in the evaluation process. The majority of the agency contacts contributed to the evaluation by providing information about the agency and/or program and any available data already collected by the agency. One stakeholder commented about having “no other presence [role] and was happy about that,” thus demonstrating how little involvement some stakeholders wanted to have in the evaluation process. Other stakeholders such as board members or external funding sources were rarely involved in any of the evaluation processes.

Overall the agencies were pleased with outcomes of their evaluations. While a few agencies utilized recommendations and made changes to program structure, the majority of agencies did not implement recommendations. Several agencies cited funding issues as a barrier to implementation. Reported lack of funds prevented programs from hiring new staff, making structural changes, or developing new instruments for evaluation or research. Others stated that the recommendations were not applicable to the program’s goal or objectives. These recommendations were not widely accepted or even thought to be genuine recommendations. Agencies that were able to implement some of the suggested recommendations did so because the recommendation fit with their agency’s capacity and/or the evaluation team provided assistance after completion of the evaluation. In one instance, a member of the evaluation team for 10%, an LGBTQ youth support agency, worked with the program director to implement a database program for the agency to collect demographic information. The program director of 10% reported that the agency would

not have been able to use the database if it were not for the evaluation team.

In determining the type of use that agencies engaged in, conceptual use was most common, occurring mostly during the process phase of the evaluation. For example, the Teachers for Tomorrow program staff immediately began thinking about how knowledge of their program filtered to students as opposed to teachers; the agency contact had never thought about Teachers for Tomorrow’s potential impacts that way. Generally, agency contacts saw the evaluation experience as a learning one, a chance to understand the agency and its missions/objectives in a different light, instead of as a true catalyst for change. This finding is not surprising given the lack of recommendations’ use combined with the overall positive report of being involved in an evaluation: All the agencies interviewed found some value in participating in a program evaluation. In one agency, the evaluation process “allowed staff to see where they are and where they need to go” while another agency contact felt “it provided a tool for ongoing development” of the program.

Assessing the time factor reveals that most agencies used their evaluations either immediately or at the end-of-cycle. Agency contacts used information gathered during the process of conducting the evaluation and found opportunities to change their program. One agency contact noted how the process of trying to find program participants to interview for the evaluation was difficult, demonstrating a problem in agency-client communication. For the Statewide Community Action Group, data needed to evaluate the program were just not there. The agency contact recognized the need to start tracking program information earlier and began seeking ways to make this during the evaluation. The United Nations program at the Boys and Girls Agency experienced similar immediate evaluation results. In the process of conducting the evaluation, the evaluation team found that only two students gave correct

pretest answers to what was supposed to be a core part of the program. The Boys and Girls Agency was so surprised to find this that it changed the program to include a guest speaker on the key topic.

Those agencies who noted end-of-cycle use drew on their completed evaluations for a month or two. Most agency contacts shared the evaluation with agency stakeholders, such as other employees, board members or outside funding sources. The sharing was generally passive: Stakeholders were given a copy of the report, and few agencies participated in active discussions about the process and/or the recommendations. The few agencies that used the evaluation more long-term did so in an active way: i-Mentor reports to refer back to the findings at almost every monthly board meeting, and 10% refers to its evaluation weekly in an effort to continue evaluation in-house and increase youth involvement.

Discussion

Consistent with previous literature is the point that most evaluation use is conceptual. This study demonstrated that few agencies actually implement any of the recommendations given in the evaluation report. Recommendations were often viewed as “nice ideas” or something to implement should there be enough resources, but never viewed as a necessity or with a sense of urgency. Interview participants often expressed how the recommendations provided ideas for change or alteration of their programs. Many interviewees discussed how the evaluation brought to light process uses in addition to the expected outcome and impact results.

The majority of the recommendations given across the interviewed agencies are classified as changes to rules and changes to the structure, as conceptualized by Downs (Johnston, 1988). These two categories of change were previously shown in the literature to suppress use. The most accepted type recommended changes in

theory, behavior changes, were not recommended often in this sample.

Some agencies reported that recommendations were unclear and accompanied by little or no direction on how to achieve them. Agency contacts expressed interest in having the evaluation team assist after the evaluation with the recommendation implementation. Mowbray (1988) noted that an important aspect of making an evaluation useful is providing time for the evaluation team to work with the agency after the evaluation is completed, assisting with the process of implementing recommendations. Our findings were consistent with this notion. That said, none of the agencies initiated requests for help, and only two of the agencies responded to follow-up offers (before this follow-up research on use) to clarify or help regarding the evaluation results.

Our results align with Taut and Alkin’s (2003) theory regarding human and evaluation factors’ influence on evaluation use. Both evaluator competence and concerns in understanding the sensitive and complex nature of the programs being evaluated were articulated. The human factors discussed frequently by the agency contacts were problems with keeping agency contacts informed of the evaluation process and progress. Agency contacts, as a whole, were pleased with the amount of contact with their evaluation team when asked directly but throughout many interviews, respondents highlighted issues of communication with the evaluation team and dissatisfaction with their own lack of input to the evaluation. As previously noted, many recommendations given to these programs were not structurally feasible or had been previously considered and rejected. Perhaps lack of communication during the completion of the report lead teams to make recommendations that never had a chance of being implemented.

Taut and Alkin (2003) further note that agency contacts and staff may be unclear on the

general benefits and potential risks of evaluation. The agency contacts in our study were quite clear about the benefits of program evaluation, and they voiced no concern of losing their job when explicitly asked about the risks of participating in evaluation. The majority of interview participants felt that even though an evaluation of their program might show the program as being ineffective or not working as intended, an evaluation was an opportunity to grow, change, and obtain outside opinions on program function. This observation may result from selection bias; that is, the agencies that sought partnership with this course to evaluate their programs may look more favorably on evaluation than might agencies being forced to undergo evaluation.

The last part of our framework considers time and the utilization of evaluation findings. Kirkhart (2000) identified three different times in which evaluation results can be used. Our findings showed a variety of time cycles for evaluation use, with a heavier concentration on the end-of-use time cycle. Many of the agency contacts identified reading the evaluation report immediately after receiving the report, sharing the report with stakeholders and other staff, and then putting the report aside because the recommendations were too difficult or complex to implement or unimportant to the program. We also observed some concurrent and long-term influence of evaluation.

Conclusion

Several themes emerge from our research. We found very little instrumental use of the evaluation findings within this sample of sixteen evaluations. We also observed that the agency contacts felt uninvolved during evaluation process, perhaps as a result of the model of third-party/external evaluation employed (that is, these were not collaborative or empowerment evaluations). Most recommendations were viewed as “ideas for change” rather than steadfast items to

implement. If instrumental use is desired, then recommendations need to be structured and make good use of program context of the program and agency capacity (Iriti, Bickel & Nelson, 2005). Connecting with agency contacts on an ongoing basis and involving them in the evaluation process appears to be one way to increase the instrumental use of program evaluation findings.

Our research finds that evaluation use is predominantly conceptual and procedural rather than instrumental and immediate. The agencies in our study did not use their evaluations instrumentally, yet found value in the evaluation process. It is clear the field of program evaluation needs to continue research into the area of evaluation use. Much of the previous literature is either theoretical or based on speculation and anecdotal experience. Our attempt here was to bring together varied theoretical perspectives and test an integrated model of evaluation use by examining a sample of small-scale evaluations in a diverse set of non-profit and public agencies. Our hope is that both our theoretical and empirical contributions will inform and motivate future research on evaluation use.

References

- Alkin, M. C. (1985). *A guide for evaluation decisions makers*. Thousand Oaks, CA: Sage.
- Alkin, M. C. (2003). Evaluation theory and practice: Insights and new directions. *New Directions for Program Evaluation*, 97, 81-90.
- Caracelli, V., & Preskill, H. (Eds.). (2000). The expanding scope of evaluation use. *New Directions for Evaluation*, 88.
- Catlaw, T. J., & L. R. Peck. (2007, October). *Preparing for governance: Creating quality MPA/MPP education through community-based learning*. Paper presented at the annual meeting of the National Association of Schools of Public Affairs and Administration (NASPAA), Seattle, WA.

- Chelimsky, E. (1983). The definition and measurement of evaluation quality as a management tool. *New Directions for Program Evaluation*, 18, 113-126.
- Chelimsky, E. (1986). What have we learned about the politics of program evaluation? *Evaluation Practice*, 8(1), 5-21.
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331-364.
- Dibella, A. (1990). The research manager's role in encouraging evaluation use. *Evaluation Practice*, 11(2), 115-119.
- Downs, A. (1967). *Inside bureaucracy*. Boston: Little Brown.
- Ginsburg, A., & Rhett, N. (2003). Building a better body of evidence: New opportunities to strengthen evaluation utilization. *American Journal of Evaluation*, 24(4), 489-498.
- Iriti, J. E., Bickel, W. E., & Nelson, C. A. (2005). Using recommendations in evaluation: A decision-making framework for evaluators. *American Journal of Evaluation*, 26(4), 464-479.
- Johnston, W. P. (1988). Increasing evaluation use: Some observations based on the results at the U.S. GAO. *New Directions for Evaluation*, 39, 75-84.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5-23.
- Leviton, L. C., & Hughes, E. F. X. (1981). Research on the utilization of evaluations. *Evaluation Review*, 5(4), 525-548.
- Marsh, D. D., & Glassick, J. M. (1988). Knowledge utilization in evaluation efforts: The role of recommendations. *Science Communication*, 9(3), 323-341.
- McLaughlin, J. A., Weber, L. J., Covert, R. W., & Ingle, R. B. (Eds.). (1988). Evaluation utilization. *New Directions for Program Evaluation*, 39.
- Mowbray, C. T. (1988). Getting the system to respond to evaluation findings. *New Directions for Evaluation*, 39, 47-58.
- Oman, R. C., & Chitwood, S. R. (1984). Management evaluation studies: Factors affecting the acceptance of recommendations. *Evaluation Review*, 8(3), 283-305.
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. *Evaluation Practice*, 18(3), 209-225.
- Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, 88, 25-37.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195-208.
- Smith, M. F. (1988). Evaluation utilization revisited. *New Directions for Evaluation*, 39, 7-19.
- Stevens, C. J., & Dial, M. (1994). What constitutes misuse? *New Directions for Program Evaluation*, 64, 3-13.
- Taut, S. M., & Alkin, M. C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation*, 24(2), 213-226.
- Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, 22(3), 387-395.
- Weiss, C. H. (1981). Measuring the use of evaluation. In J. A. Ciardo (Ed.), *Utilizing evaluation: Concepts and measurement techniques* (pp. 17-33). Beverly Hills, CA: Sage.

Appendix A: Interview Questions for “Evaluating the Evaluations” Project

In general...

1. Please describe, in your own words, the definition of “program evaluation.”
2. Please describe some of the benefits of having your agency participate in a program evaluation. Please describe some of the possible consequences.

Before the evaluation...

3. Before you had participated in this evaluation in the spring of (insert date)...
 - a. What was your or your agency/organization’s prior experience with program evaluation?
 - b. What was your main goal or expectation of becoming involved in this program evaluation project?

4. How much input/time/resources did you expect to provide during the evaluation process?

During the evaluation...

5. Please describe any other experiences you or your agency had with graduate students (e.g. interns).

For questions 6-10: During the evaluation process in spring of (insert year) a student evaluation team conducted an evaluation of your program.

6. After an initial meeting (or direct contact) with your student evaluation team, how confident were you of the students’

competence in conducting the program evaluation? If you had any concerns, please describe.

7. Describe your level of participation during the evaluation process (e.g. data collection, providing feedback).
8. How often were you in contact with your student evaluation team or the professor? Were you satisfied with this amount of contact? Why or why not?

Reflective...

9. During the process of conducting the program evaluation, did anything change your way of thinking about your program and/or the agency as a whole? Did you take any action on these thoughts?
10. Please describe your overall relationship with the student evaluation team or the professor. Were you satisfied with this relationship?

After the evaluation...

For questions 11+: In May of (insert year), you received a final, written report of the evaluation conducted by the student evaluation team

11. Please describe your initial thoughts after reading the evaluation report.
12. How well do you feel the student evaluation team performed an accurate evaluation of your program? Please describe elements of the evaluation that you did/did not agree with (not including the final overall results). For example, were you in agreement about the scientific validity of the report, choice of data collection, etc.?
13. In your agency’s program evaluation report, the following recommendations were given:

(insert specific recommendations from each report). Please comment on the implementation or the lack of implementation for each recommendation. (For each recommendation) Why or why not was this recommendation implemented?

14. Please describe any barriers outside your direct agency that might have made it difficult to use the recommendations given in the program evaluation (e.g. political, funding).

a. Were you contacted at all by your student evaluation team after receiving the evaluation? Did you contact your student evaluation team after receiving the evaluation? Why (or why not)?

b. Were you contacted at all by the professor after receiving the evaluation? Did you contact the professor after receiving the evaluation? Why (or why not)?

Reflective...

15. Do you feel the evaluation completed by the student evaluation team was well done (e.g., did the design and/or the analysis accurately reflect the program?)? Why or why not?

16. How well do you feel the student evaluation team got at the crux of the important questions within your agency/program?

17. Please describe an aspect of the evaluation process that was unexpected. Please describe an aspect of the evaluation results that was unexpected.

18. In the first six months after student evaluation team completed the evaluation of your agency/program, how often did you refer to the evaluation?

a. Overall since the completion of the evaluation of your agency/program, how often do you refer to the evaluation?

19. Have you shared the evaluation with anyone else, including coworkers, stakeholders, board of directors, or funding sources? Why or why not? If so, who? And what were their responses?

20. Overall, were you pleased with the outcome of the program evaluation? Why or why not?

a. Were you pleased with the experience of having participated in an evaluation of your program? Why or why not?

21. Have you participated in an evaluation of your program again?

a. If so, was the evaluation for the same or different part of the program/organization that the student team worked on? Please describe.

b. If not, would you participate in an evaluation of your program again? Why or why not?

Appendix B: Survey

“Evaluating the Evaluations” Project
Professor Laura R. Peck, ASU School of Public Affairs, with Lindsey Gorzalski
Survey for Agency Contacts/Participants in Program Evaluation
August 29, 2005

Instructions: Please respond to each of the following questions by circling the appropriate number. There is no right or wrong answer; instead we are asking for your opinion about evaluation in general and for your experience working with ASU’s PAF 541-Program Evaluation course in conducting evaluation research in your agency. Your answers are strictly confidential and will not be shared with anyone in your agency.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Program evaluation is important to my agency.	1	2	3	4	5
2. Program evaluation can tell me whether or not a program is working as intended.	1	2	3	4	5
3. Participation of an agency or organization is critical in conducting a program evaluation.	1	2	3	4	5
4. The process of participating in a program evaluation affected my thoughts and feelings about my agency or program.	1	2	3	4	5
5. It was easy for me to work with the student evaluation team.	1	2	3	4	5
6. I found the student evaluation team to be knowledgeable about program evaluation.	1	2	3	4	5
7. I was confident in the competence of the student evaluation team.	1	2	3	4	5
8. I was confident in the competence of the professor of the program evaluation course.	1	2	3	4	5
9. The findings of the evaluation report were favorable to my agency.	1	2	3	4	5
10. The recommendations given in the evaluation report were utilized.	1	2	3	4	5
11. I was satisfied with the report produced by the student evaluation team.	1	2	3	4	5
12. The evaluation report was useful to my agency.	1	2	3	4	5

(_____ / _____)