April 2021

Relationship Between Working with Professional Evaluators and an Organization's Evaluation Culture

James M. Wharton
University of South Florida

Relationship Between Working with Professional Evaluators and an Organization’s Evaluation Culture

by

James M. Wharton

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Curriculum and Instruction
with a concentration in Measurement and Evaluation
Department of Educational and Psychological Studies
College of Education
University of South Florida

Co-Major Professor: Liliana Rodriguez-Campos, Ph.D.
Co-Major Professor: Robert F. Dedrick, Ph.D.
John M. Ferron, Ph.D.
Waynne B. James, Ed.D.

Date of Approval:
March 31, 2021

Keywords: zoos, aquariums, evaluative thinking, psychological safety

Acknowledgments

I'd like to thank and acknowledge my committee for their guidance and support: co-major professors, Drs. Liliana Rodriguez-Campos and Robert Dedrick and committee members Drs. John Ferron and Waynne James. Dr. Rodriguez-Campos has been my primary mentor in my graduate-level evaluation learning. Her work in collaborative evaluation and my experience in her course and working together on the case-study volume have been a significant influence on my approach to evaluation (and as a people manager). Dr. Rodriguez-Campos also encouraged me to continue when the pressures of distance and a fulltime career had me ready to throw in the towel. Dr. Dedrick's guidance on the methodological elements of my project has been invaluable. His challenge to finish by the end of 2020 was likely the final push I needed to complete my studies (though I didn't quite meet that marker). Despite my remote location and protracted candidacy, all members of the committee have been responsive and helpful through the entire process.

I would also like to thank Kathayoon Khalil, Joy Kubarek, Marley Steele-Inama, Kelly Riedinger, and John Fraser for their expertise, advice, and encouragement; Bob Davidson, Brad Rutherford, and the staff and leadership of the Conservation Engagement and Learning Department for their support and patience; Seattle Aquarium and Mote Marine Laboratory for financial support, and my family for enduring this long, drawn-out process and my distraction throughout.

Table of Contents

List of Tables

List of Figures

Abstract

Mission-based organizations like zoos and aquariums are investing in evaluation capacity to help them improve their mission performance, but are these resources improving their professional culture, or merely creating evaluations? This study surveyed the leadership from 100 programming departments at accredited U.S. zoos and aquariums to learn how work with professional evaluators might be related to the nature of an organization's evaluation culture. Survey results showed no statistically significant relationships between a self-reported measure of evaluation culture and either institutional demographics or work with professional evaluators. Follow-up interviews with nine case study organizations, however, were more supportive of the role evaluators play in improving practice, suggesting management or structural limitations may be limiting impact. An exploratory factor analysis revealed the emergent construct of psychological safety as a potential new avenue for future research around the antecedents of evaluation culture. Reflections from case study interviews and existing literature are synthesized to offer recommendation for professional practice.

Chapter One: Introduction

Zoos and aquariums are moving from tourist attractions with educational benefits to conservation organizations. Ogden and Heimlich (2009) succinctly summarized this evolution of the modern zoo and aquarium (see also Brewer, 2001; Fraser & Wharton, 2007; Rabb, 2004). The Seattle Aquarium is an example: The Aquarium was once a city-owned community attraction and resource. After transitioning to non-profit management, the Aquarium adopted a conservation mission (Inspiring Conservation of our Marine Environment) and created a strategic plan that recast the institution as a conservation organization with education and community-building strategies (Seattle Aquarium, 2011). An organization's mission defines its purpose and the ultimate outcomes against which it has chosen to be measured (although organizations can and do generate other outcomes through their activities, and external audiences may evaluate organizations with by their own outcomes). Using the Seattle Aquarium as an example again, while its mission is most directly addressed through conservation outcomes, it also creates educational and entertainment-based outcomes as guests learn and enjoy themselves as a result of a visit. The city of Seattle may also expect the Aquarium to create economic impact on the Waterfront or community value by serving under-resourced audiences.

With this new emphasis, zoos and aquariums are insisting on credible metrics and measures to help them understand how they can achieve greater mission impact. To develop these metrics and understand their impacts, zoos and aquariums may work with external evaluators, audience research consultants, hire internal expert evaluation staff, or rely on the professional experience and effort of existing staff.

I began my studies hoping to better understand how to measure the mission impact of my organization. At the beginning, I was the Vice-President for Education at Mote Marine Laboratory and Aquarium and in the course of my degree, I accepted a position as Director of Conservation Engagement and Learning at the Seattle Aquarium. My leadership role put me in a position to consider how we could demonstrate and improve our mission performance. Evaluation was clearly a tool at our disposal. While my initial idea was to leave with practical methods for evaluating mission outcomes, my professional experience suggested evaluation efforts in isolation could be ephemeral and unsustainable. Grant-funded external evaluations, or even purposeful internal evaluation efforts face challenges for meaningful use (Khalil & Ardoin, 2011; Roe, Mcconney, & Mansfield, 2014). These evaluations may *sit on the shelf* and fail to make substantive changes and/or improvements in institutional programming. Internal evaluation efforts struggle to maintain priority in the crush to do more programs, serve bigger audiences, and generate more revenue (Clavijo, Fleming, Hoermann, Toal, & Johnson, 2005; Luebke & Grajal, 2011; Monroe et al., 2005; Roe et al., 2014).

Fostering an organizational context that values the learning generated by evaluation is one way to ensure an institution's investment in evaluators is productive. An organization which invests in and values the idea of using evaluative thinking for program improvement and more effective decision-making might be said to have a strong *evaluation culture*. Others refer to these as learning organizations. Hiring internal expert evaluation staff is one way an organization signals that it values this kind of evaluation culture, but it may also result from the development or refinement of a weak or nascent culture within the institution. Hiring internal staff represents costs to the organization and requires the prioritization of limited resources for this evaluative function.

Problem

While there is established scholarship around the process benefits of participating in evaluations and around the varied roles of evaluators (that can include teacher and coach), what is not clear from the literature is the relationship between work with professional evaluators and strength of an institution's evaluation culture. Institutions are investing in evaluation resources—internal, external, and through professional development—but is this investment enhancing evaluative thinking within the organization, or merely generating evaluations?

Purpose

The purpose of the research is to survey the leadership from programming departments at accredited U.S. zoos and aquariums to learn how working with professional evaluators might be related to the nature of the evaluation culture at these mission-based organizations.

Rationale

In community-serving non-governmental organizations like zoos and aquariums, the activities an institution undertakes to address its mission are sometimes referred to as the organization's *program*. If this mission-based work can be collectively referred to as program, then mission performance is improved through program improvement. An organization, therefore, should strive for a culture where management supports staff who regularly engage in reflective, improvement-oriented practices which systematically collect/use data to make context-appropriate decisions, including engaging in formal evaluation activities as warranted. This can be described as a strong *evaluation culture*. One way to develop evaluative thinking or an evaluation culture with staff is to work with evaluators. This work might consist of identifying program outcomes, creating evaluation questions, designing or reviewing data collection instruments, collecting or analyzing data, and/or discussing conclusions and recommendations. Increasingly, these process benefits are valued as highly as the instrumental benefits provided by the findings of any individual evaluation (Hargreaves & Podems, 2012; Patton, 1998; Preskill & Zuckerman, 2003).

Some organizations have invested in internal evaluation capacity (one or more staff whose primary responsibility is evaluation and who possess some professional experience or training), while others work exclusively with external evaluators, or not at all. Understanding whether it is beneficial to an organization's evaluation culture (and thereby its mission performance) to employ internal evaluation capacity is important, because these represent a strategic use of limited resources that could be employed for program development or implementation.

Research Questions

How does the evaluation culture of an organization vary related to its work with professional evaluators? Associated questions include:

- Are organizations with internal evaluation staff associated with the strongest evaluation cultures?
- Are organizations that have chosen not to work with professional evaluators in any capacity associated with the weakest evaluation cultures?

- How do organizations compare that work with a combination of internal and external evaluation staff?

Institutional demographics, including the form or governance and the size of an institution could influence how an organization may, or may not, work with evaluators. They may also influence some aspects of an organization's evaluation culture. Therefore, secondary questions around institutional demographics will be explored, including:

- Does work with professional evaluators vary with institutional demographics (form of governance, annual operating budget, annual attendance)?
- Does an organization's evaluation culture vary with its institutional demographics?

Hypotheses

Organizations that have invested in internal evaluation staff would be associated with the strongest evaluation cultures (as assessed in this study). Organizations that work with a combination of internal and external evaluators may be associated with slightly stronger evaluation cultures, but the differences would not be significant in survey results.

Delimitations

In studying zoos and aquariums, this study limited consideration to institutions accredited by the Association of Zoos and Aquariums (AZA). AZA is a national and international professional society and accreditation organization that certifies member institutions that meet the highest professional and community standards for animal care, educational merit, conservation efficacy, and financial sustainability. AZA accreditation standards require education and interpretive programs be evaluated "on a regular basis for effectiveness and content," looking for both participant satisfaction and program impact. AZA accreditation standards specify that evaluation results should be used for improvement (Association of Zoos & Aquariums, 2019, p. 22). This standard may be met through the activities of internal, non-evaluation staff and/or may be identified as an area for improvement even as the institution is granted accreditation on the overall merits of its application.

This study included only U.S.-based institutions as these were less likely to demonstrate significant cultural differences (including legal, ethnic, and language differences). Participation in a professional society that requires accreditation as a condition of membership further reduced potentially complicating cultural variability. As of October 2018, there were 233 accredited institutions with 215 located in the U.S. (Association of Zoos & Aquariums, 2018b)

Within these institutions, this study focused on working groups and departments that create educational and interpretive programming for visitors and community members. When evaluation staff are present in AZA institutions, they are very often part of these departments. It is also in these departments where an effect from working with evaluators is likely to be seen. Within these departments, the primary informants were the top-level managers (typically a *director* or *vice-president*). While it is certainly feasible, even expected, that aspects of an evaluation culture could spread to other departments within the organization (finance, administration, etc.), that is beyond the scope of this study.

Limitations

Focusing on zoos and aquariums may limit the generalizability of findings to other mission-based organizations. Zoos and aquariums are unusual in the realm of either conservation or social service non-profits in that they provide a variety of business and social services as visitor-serving institutions with conservation missions. Some aspects of zoos and aquariums may be as revenue-driven and business-oriented as any tourist-serving for-profit company, while other aspects provide social services in the form of community outreach, serve an educational function with schools and families, or conduct scientific research locally and globally.

All survey and interview data were self-reported and voluntary. There is always the potential for bias in who chooses to respond to the invitation. Those with established interest in or experience with evaluation may have been more likely to participate, skewing results positively. An evaluation culture score and indeed all staff assessments of the culture within a workgroup or organization are only a snapshot judgement of a subset of staff within an institution. Focusing on leadership and the programming departments may have further limited the generalizability of findings. This study did not attempt to establish a causal relationship between the presence of internal evaluation staff and the development of an evaluation culture within the organization.

It should also be noted that this study was conducted in an unusual working climate: a global pandemic that resulted in wide-spread zoo and aquarium closures and lay-offs. Education and programming departments were particularly hard hit. Animals need to be fed and cared for no matter the circumstances, but with no guests and extensive community restrictions, programming activities were significantly reduced. Respondents were asked to answer survey questions considering their pre-COVID conditions, but as questions contain references to organizational culture, leadership, systems, and teams, it would be hard to imagine that respondents were not influenced by their circumstances at the time of testing.

Definition of Terms

The **Association of Zoos and Aquariums** *(AZA)* is "a 501(c)3 non-profit organization dedicated to the advancement of zoos and aquariums in the areas of conservation, education, science, and recreation. AZA represents more than 240 facilities in the United States and overseas, which collectively draw more than 200 million visitors every year" (Association of Zoos & Aquariums, n.d.). Zoos, aquariums, and other conservation institutions that care for live animals can be accredited by AZA. **Accreditation** is an exhaustive application and review process that assesses animal welfare, staff and visitor safety, educational merit, conservation efficacy, and financial sustainability against the highest professional standards. Institutions may apply for accreditation once every five years.

Audience researchers are a category of evaluator in this study. Audience researchers develop market research studies to understand the available visitorship within a community. They may also survey or otherwise inquire about the quality and value of the visitor experience among guests to a zoos or aquarium (or museum).

The units of analyses in this study are the education/engagement **departments** within U.S.-based AZA-accredited zoos and aquariums. Because the organizational charts of zoos and aquariums sometimes subsume the education/engagement activities and staff into another department (e.g., Conservation), the terms department and **work/working group** will be used interchangeably.

Director or **education director** refers to the senior department manager. Depending on the organization and their hierarchy, that role may be referred to as a director, curator, or vice-president.

The education departments of some zoos and aquariums have begun to replace the word *education* in their titles with **engagement**, recognizing that education is a strategy towards the ultimate goal of these departments in conservation-oriented zoos and aquariums, to get visitors and communities engaged in conservation.

An **evaluation culture** is one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions—with the mandate and support of organizational leadership.

Overall **evaluation culture scores** in this study have been generated by averaging the mean scores from the six constituent subscales (or **dimensions**, used interchangeably) from the modified Readiness for Organizational Learning and Evaluation (ROLE) survey instrument.

Evaluative thinking is a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making.

An **external evaluator** is a professional evaluator (see definition below) working with an organization on a contract or project basis. These evaluators could include program evaluators, audience researchers, or university partners. External evaluators may provide their services for a fee or on a pro-bono basis.

Institutional governance in this study refers to the operating authority of a zoo or aquarium. Some are operated by for-profit companies, some by non-profit organization, and still others are publicly operated by a government agency or municipality.

Institutional size in this study is defined through two categories, operating budget and annual attendance. These categories are defined by AZA and reported on through annual benchmarking surveys. See Table 4 for details on how small, medium, medium-large, and large are defined for each category.

An **internal evaluator** is a professional evaluator (see definition below) that is employed as a member of the staff of an organization. Internal evaluators may work full or part time and may serve as program evaluators, audience or social science researchers, or work across both evaluative and research functions.

A **professional evaluator** is one with formal training/education in evaluation (e.g., at a university or through a professional certification program) and/or several years of work experience in an evaluative function.

Program refers to the collective activities undertaken by a non-profit organization to meet its mission.

Trained (non-evaluator) staff include paid internal staff of an organization with education or experience comparable to a professional evaluator (see definition above), but for whom evaluation is not a primary job function (an educator or manager, for example).

Chapter Two: Literature Review

As mission-based conservation organizations, zoos and aquariums are under pressure to demonstrate their effectiveness. Some of this pressure is external. Private and federal funders, even institutional members are increasingly interested in seeing a demonstrable return on their investment. Animal rights groups are fomenting public pressure for institutions with animals in human care to justify the imposition on these animals' lives. However, as much or more pressure comes from internal drivers, namely, the professionalization of zoo and aquarium staffs and an interest in continuous learning and improvement in service of their conservation missions. This has led to working with evaluators and evaluation staff. To answer the question of the relationship between evaluation staff and the culture of an organization or workgroup, this chapter will explore the context of evaluation in the zoo and aquarium setting, the strengths and weaknesses of current evaluation capacity-building models, the development of evaluative thinking in professional staff, the current understanding of what it means for an institutions to demonstrate an evaluation culture, and the differences between working with internal and external evaluation staff.

Professionalization of Staff

Zoo and aquarium educators believe they know their activities are impactful intuitively; they see it in the reactions and language of their participants, but documenting the impact in a scientifically rigorous way has not always occurred. When it has, it has primarily been through large, well-resourced institutions. Working from this assumption of effect, zoos and aquariums have measured their impact with largely effort-related metrics (e.g., annual attendance, number of schools and schoolchildren served, number of teachers trained) and satisfaction surveys (Monroe et al., 2005).

This is changing as institutions and individuals in the field of zoos and aquariums are becoming more interested in the meaningful evaluation of program outcomes (Heimlich & Horr, 2010; Khalil & Ardoin, 2011; Kubarek & Trainer, 2015; Somers, 2005). Though the educational preparations, experiences, and credentials of zoo and aquarium educators are likely to be as diverse as those of science museum educators, common funders and professional networks (e.g., AZA, North American Association of Environmental Educators, National Science Teachers Association, National Marine Educators Association, National Association for Interpretation) have contributed both to the professionalism of the field (Uyen Tran & King, 2007) and elevated the importance of evaluation (Khalil & Ardoin, 2011). As one example, the Conservation Education Committee of AZA has recently developed a framework for social science research in zoos and aquariums to help direct and coordinate the growing scholarship in this context (Fraser et al., 2010). While the establishment of zoo and aquarium education as a profession may not be imminent, the increased interest in creating and improving upon effective practices is evident.

Evaluation in Zoos and Aquariums

To understand evaluation culture-building in zoos and aquariums, we must understand the past and present of evaluation efforts in the field, the challenges facing improved evaluation practice, and what efforts have been previously applied to addressing the problem.

The evolution of evaluation in zoos and aquariums. As noted earlier, zoos and aquariums have only begun to think of themselves as education or conservation organizations in the last several decades. As such, evaluation for these newer outcomes has been underdeveloped. Instead, success has been viewed through a combination of effort metrics (e.g., program participation, visitation) and satisfaction surveys (Luebke & Grajal, 2011; Roe et al., 2014). Over time, evaluation outcomes diversified to include program effectiveness, improvement, increased funding, expanded engagement with stakeholders. Even professional peer pressure played a role (Khalil & Ardoin, 2011; Roe et al., 2014). While the motivations for evaluation can be myriad and well-meaning, Stufflebeam and Shinkfield (2007) would caution that some motivations, like public relations-inspired studies, could stray into the realm of pseudoevalution if evaluation standards are not strictly adhered to.

Early efforts focused on cognitive outcomes (Ogden & Heimlich, 2009). As the focus of education programs expanded to include affective and behavior change outcomes, evaluation efforts to match lagged (Fraser & Sickler, 2009; Luebke & Grajal, 2011). Two recent studies looked at the state of program evaluation (and, to a lesser degree, visitor research) in zoos and aquariums domestically and globally. Luebke and Grajal (2011) surveyed 97 AZA-accredited zoos and aquariums to better understand the extent to which each was measuring mission outcomes. The study found that although most zoos and aquariums had conservation missions, most learning outcomes associated with programming reflected cognitive outcomes and most evaluation efforts were focused on demographics (including extensive marketing surveys) and participant satisfaction. The researchers also found, although most institutions wanted to improve their evaluation efforts, they faced a consistent set of barriers including money, resources, staff time, and internal expertise (especially regarding methods to measure affective and behavior change outcomes). A second study by Roe et al. (2014) collected survey information from 176 institutions with nine follow-up case studies from zoos and aquariums across the globe. The researchers found most respondents had a basic understanding of evaluation. The most common motivation expressed was for program improvement, although the methods used for evaluation did not always match this goal (e.g., satisfaction surveys and participation metrics). Evaluation efforts consisted most commonly of setting program objectives (outcomes), less often measuring achievement of objectives, and even less often using the results. The researchers cited a very similar set of institutional barriers to growing evaluation efforts, primarily time, knowledge, and money.

This pattern of perceived barriers to increased evaluation efforts arises consistently from these and other studies (Clavijo et al., 2005; Khalil & Ardoin, 2011; Ogden & Heimlich, 2009). They include obvious hurdles like time, funding, and expertise, but also include more systemic challenges like limited interventions with audiences, seasonality of programming, and a strong reliance on external funding (Clavijo et al., 2005). Any model of evaluation capacity building cannot only address the knowledge of and attitudes towards evaluation use, it must also address the perceived barriers which prevent more widespread adoption of evaluation activities, but are the existing models of evaluation capacity building more focused on creating evaluators rather than evaluation?

Evaluation Capacity Building

Increasing evaluation capacity is not a challenge facing zoos and aquariums alone. There is a body of literature which addresses evaluation capacity building (ECB) in organizations (and a smaller body which addresses zoos and aquariums specifically). Definitions of ECB vary throughout the literature. Hueftle Stockdill, Baizerman, and Compton (2002) offer both conceptual and working definitions for ECB in their literature review which introduces a special edition of the journal *New Directions in Evaluation* dedicated to ECB. Their conceptual definition emphasizes the contextual nature of ECB, its intentionality, and its sustentation, while the working definition focuses on primarily the latter two concepts. Labin, Duffy, Meyers, Wandersman, and Lesesne (2012) provide a definition which emphasizes intentionality, evaluation competency, and the use of results, but also discusses the importance of context in the form of organizational support and culture.

The idea that ECB is intentional is consistent throughout the literature. Although conducting and/or participating in evaluations may help develop evaluation competencies and improve attitudes toward evaluation (Labin et al., 2012; Monroe et al., 2005; Suárez-Herrera, Springett, & Kagan, 2009), one cannot assume that merely the act of participation will result in capacity building outcomes. Therefore, like other educational activities, ECB efforts should have clear goals and objectives, be rooted in adult learning theory, and understand the organizational and cultural contexts of the learners (Preskill, 2008). Also consistent is the idea that ECB should strive to create an evaluation program which is sustainable, routine, and integrated into institutional practice. This is sometimes characterized as synonymous with or highly influenced by organizational culture.

Hueftle Stockdill et al. (2002) characterize ECB as the intersection of the overall evaluation process, actual evaluation practices, and the occupational orientation/practitioner role, all oriented to organizational structures, culture, and broader workplace practices. This orientation brings common barriers to ECB to the surface, including the need for broad stakeholder support, dedicated resources, and a clear understanding of the outcomes of ECB and how they relate to an institution's program or mission outcomes. The ultimate outcome of ECB should not only be to increase the capability of staff to conduct evaluations, but also to create improved program/mission outcomes (Wandersman, 2014).

ECB is often characterized as a multi-modal, mixed methods approach (Cousins, Goh, Elliott, & Bourgeois, 2014; Hueftle Stockdill et al., 2002; Labin et al., 2012; Preskill, 2008; Suárez-Herrera et al., 2009). Cousins et al. (2014) defines ECB activities as either direct or indirect. Direct activities are more instructional and intended to build knowledge and skill. Trainings, college courses, and other workplace professional development activities are examples. Alternatively, indirect activities are more experiential. This is learning which occurs while conducting, assisting, or participating in evaluation work. Collaborative, participatory, and empowerment evaluation approaches are especially well suited for providing this kind of learning (Labin et al., 2012; Monroe et al., 2005; Suárez-Herrera et al., 2009). This is also referred to as process use (Patton, 1998). Indirect activities also help build positive affect and attitudes toward evaluation practice, outcomes which are underemphasized in many ECB efforts (Labin et al., 2012).

While the ultimate outcome should be to improve program or institutional mission outcomes, direct outcomes associated with ECB exist at both the individual and institutional level. For individuals, ECB outcomes include improvement in evaluation knowledge/skill and in attitudes towards evaluation generally. For organizations, ECB should result in the establishment of evaluation processes, policies, and practices, increased resources available for evaluation, improved support from leadership and within the organizational culture, all leading to a mainstreaming, or normalization of evaluation within the organization. These outcomes are dynamically related. Without sufficient evaluation knowledge and positive attitudes towards evaluation among the staff, some organizational outcomes (e.g., support from leadership, increased resources, change in organizational culture) may be difficult to achieve. On the other hand, without strong leadership willing to dedicate resources, building evaluation competencies which support continuous improvement of mission-based activities will also be a challenge. Labin et al. (2012) suggest there may be a sequence of some ECB outcomes (e.g., changes to processes, policies, and practices may necessarily precede mainstreaming)and that variables like culture, leadership, and resources might be more like readiness factors. A challenge with these and other existing ECB models is that they seem to be biased towards creating evaluators and mainstreaming evaluation into organizational practices. Learning is centered around building evaluation knowledge and skills for individuals. For organizations, the priorities rest on securing resources and creating more support for the use of evaluation.

> In my view, (mainstreaming evaluation) implies that we are trying once again to put evaluation at center stage. As evaluators, we need to think through the extent to which our desire to mainstream evaluation is an attempt to grow the profession, in contrast to simply getting people to be more evaluative.
>
> (Duignan, 2003, p.12)

Even if successful, with dedicated resources and staff, the benefits of evaluation and evaluative thinking run the risk of being isolated in programming activities, creating a siloed pool of special expertise that does not have the impact on the operation of an organization that it could.

Evaluation capacity building efforts in zoos and aquariums. While evaluation is still a relatively new area of emphasis for zoos and aquariums, there have been several efforts at building evaluation capacity within and across organizations, recognizing the persistent challenges of time, resources, expertise, and motivation. One way zoos and aquariums have attempted to build evaluation capacity is through participation in evaluation projects. Somers (2005) describes an example in the evaluation of the Denver Zoo's Wonders in Nature—Wonders in Neighborhoods program (WIN). The WIN evaluation took a participatory approach, an approach particularly well-suited to informal education programs which are inherently participatory by design. A stakeholder-centric approach like this requires involvement from staff at all stages, providing ample opportunities to learn about evaluation design, methods, analysis, and use. While participatory evaluations can be more expensive because they require more time for stakeholder engagement, they are cost efficient as an ECB strategy because they mitigate the need for additional formal training. In this example, ECB was a named outcome of the process from the beginning, but this may not always be the case when organizations use the opportunity provided by a required evaluation for ECB. As has been established, effective ECB efforts must be intentional and systemic. There is some risk that learning from a participatory evaluation process may not be internalized (or institutionalized) if there are not additional opportunities to apply new learning (i.e., an intentional strategy to support the transfer of learning).

Participatory approaches can provide effective ECB for organizations where there is some established interest or priority for evaluation, but where resources and internal expertise are lacking. Using an external evaluator as a coach, especially one who has a history with the organization, can address the absence of expertise. Another approach is described by Owen and Visscher (2015). The authors describe an ongoing partnership between the University of Washington's Museology program and the local informal education community (including the Seattle Aquarium). The Museology program offers a two-year audience research specialization which provides graduate students training in evaluation and practical experience at museums in the community. Museums provide project ideas and connect with a team of students in their second year of the program. After a shared orientation and goal-setting workshop, students work on-site with staff to create an evaluation proposal. In the next semester, the students conduct their evaluation (with the help of first-year students) and share their results in a presentation to museum staff. The program has been a successful partnership which provides (external) evaluation capacity to museums and real-world experiences for students. Additionally, for the staff immediately associated with the evaluated program, there is an opportunity for evaluation knowledge- and skill-building as they work with the students in the goal-setting workshop, work with them to develop evaluation questions, to facilitate the students' activities, and to collect data. Although staff are not likely to move from this experience to developing and conducting their own evaluations, it is important to build skills in working with evaluators as well as working as evaluators. Understanding how to frame a program for an evaluator (including the establishment of program outcomes, and the development of logic models), understanding how to create or recognize realistic evaluation questions, and introducing them to relevant facets of organizational culture and procedures are all critical to the ultimate success of any evaluation. Developing these skills in an organization will save time and money in the end as they prevent an evaluator from frustration and rework. The Seattle Aquarium has extended the ECB opportunity further by requesting staff be included with first-year museology students in training and data collection. This has been successful as it provides additional capacity to the students to collect more and better data and has created interest and enthusiasm for evaluation practice among Aquarium staff. Participating in evaluation studies not only builds skills but also demystifies the process, reducing evaluation-related anxieties.

The evaluation activities at Shedd Aquarium have gone a step beyond participatory ECB. Shedd has invested institutional resources in evaluation staff. There is risk, however, even in this approach. With dedicated evaluation staff there is the possibility that program staff see evaluation as no longer their responsibility. Shedd addressed this risk by adopting an empowerment approach to their evaluation efforts, including the development of an evaluation toolkit that contains "vetted instruments, evaluation techniques, operationalized approaches," and the training to use them (Kubarek, 2015, p. 10). With an empowerment approach, the evaluator acts as a trainer and *critical friend* for staff as they conduct their own evaluation. With this approach, Kubarek notes, ECB is as important an outcome as the evaluation itself. The toolkit was an important element in their process as it provided an opportunity for learning as staff were trained in its use. The toolkit also provided a consistent evaluation approach throughout the organization. Having been developed by evaluation staff and designed with evidence-based practices and theory in mind, the toolkit gave evaluation efforts credibility, even when conducted by program staff. Similar to other immersive approaches (e.g., Arnold, 2006), the evaluator then can provide technical assistance as staff develop their knowledge and skills through the practical application of the toolkit. The ready access to this kind of counsel allows the staff to be flexible in the application of the toolkit without worrying about invalidating the results. Through the consistent application of the toolkit and with the availability of technical assistance from the evaluation staff, Shedd's goal is to engender a culture of evaluation which features evidence-based decision-making, cycles of reflection and action, and a community of learners. As has been noted repeatedly, the authors not only stress the importance of institutional support and investment, but also the importance of developing strong communication and coaching skills among the evaluation staff (which should not be assumed).

The final two examples of ECB in zoos and aquariums from the literature adapt established ECB models to their distinct situations. Matiasek and Luebke (2014) describe ECB efforts at the Brookfield Zoo (Chicago Zoological Society,) but make special point to extend the goals of evaluation from program improvement to organizational (mission) success. They describe evaluation capacity as achieved when:

> educators understand how their programs contribute to the organizational mission, are able to define relevant program goals, align program elements and activities to program goals, and feel confident in developing program performance measures that are consistent with their program and the organization's stated goals. (Matiasek & Luebke, 2014, p. 78)

The Chicago Zoological Society's approach draws from mission-focused, participatory, theory-based, and utilization-based approaches. Their process begins with logic models tying program outcomes to mission outcomes. Evaluation staff work from these outcomes to develop indicators and instruments tailored to the program. All information is collected online, and the results are reviewed with staff annually. The authors found the approach created ownership of the evaluation process among staff and led to more evidence-based decision-making. Like the Shedd approach (Kubarek, 2015), the authors highlight flexibility, staff engagement, and management support as keys to success in building and evaluation culture.

In this final example, Steele-Inama (2015) describes an ECB process undertaken by a community of informal education institutions. The Denver Evaluation Network (DEN) arose from discussions among colleagues about the potential for leveraging the evaluation capacity of a few well-resourced organizations to create capacity throughout a network. The network soon evolved from a community of practice to investigating collective impact by collecting and analyzing data across institutions. Through a grant from the Institute for Museum and Library Services, DEN formalized their work to assist other networks of community-based organizations. The outcomes of the project were to build the evaluation capacity of community-based organizations throughout the Mountain West, to disseminate the method of DEN, and to create an evaluation toolkit to provide practical assistance for resource challenged institutions. The DEN approach drew directly from the Multidisciplinary Model of ECB (Preskill & Boyle, 2008) through training, technical assistance, meetings to develop a community of practice, participatory learning, an evaluation toolkit, and an online portal to organize and share their results. Recognizing the oft-cited importance of organizational leadership buy-in, the Network created an annual breakfast for the Chief Executive Officers of Network members where the work and achievements of the preceding year were presented and celebrated. Keys to success shared by the authors echo familiar themes, especially flexibility and leadership support. However, being a collective effort of multiple organizations, the authors also stressed effective partnership practices like memoranda of understanding, finding the right partners, clear outcomes and timelines, and a strong core cross-institutional leadership.

Developing Evaluative Thinking

In many of the ECB approaches and examples mentioned above, learning is centered around evaluation purposes and skills, and affective development around positive regard for the value of evaluation to mission and organizational success. Following these processes may effectively result in more evaluations conducted, but will it have the desired impact on the organization's function and outcomes? Is it more valuable to teach staff how to conduct evaluations, or how to think evaluatively? Patton (2008) claims, "this kind of thinking can have more enduring value than a delimited set of findings. . . [s]pecific findings typically have a small window of relevance. In contrast, learning to think and act evaluatively can have an ongoing impact" (p. 153).

Evaluative thinking has a range of definitions in the literature, though there is at least some concern the term has been used more like a catch phrase than an academic construct (Vo, Schreiber, & Martin, 2018). Buckley, Archibald, Hargraves, and Trochim (2015) characterize evaluative thinking as critical thinking in an evaluation context. Patton (2018) would agree but goes further, suggesting inferential, creative, and practical thinking skills are also essential. Vo and Archibald (2018) also suggest evaluative thinking is like critical thinking, except that a judgement of value is required. Though Schwandt (2018) suggests that finding the perfect consensus definition of evaluative thinking is not necessary for effective learning and practice on the construct, it is helpful to look at the elements most commonly associated with the concept in the literature so that a definition might be chosen for the context of this study which is representative of previous scholarship. Table 1 describes the elements most commonly associated with definitions of evaluative thinking from the literature. While many of these concepts have reasonable analogs in the broader concept of critical thinking, inclusion of value judgements and the social nature of learning in an evaluative context begin to create some distinction. The definition below encompasses the key concepts from the literature, with the understanding that decision-making is inclusive of traditional evaluative decisions on the merit, worth, or value of a subject. It would also be reasonable to subsume the identification of assumptions and positionality into the idea of a reflective process, but it felt important to call out a practice that is commonly overlooked. This inclusion may make the definition more robust to community and workplace cultural changes spurred by events in 2020, for example, that have accelerated conversation (and hopefully progress) associated with race and social justice.

> *Evaluative thinking is a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making.*

The definition is similar to Baker and Bruner (2006) with the addition of social learning, the identification of assumptions and positionality, and the idea of organizational context appropriateness.

Table 1

Elemental Concepts Associated with Evaluative Thinking from the Literature

Concepts	Description	Sources
1. Reflective practices	Including dialogue, asking thoughtful questions, openness to change, personal accountability	(Baker & Bruner, 2006; Buckley et al., 2015; Fierro et al., 2018; Preskill & Torres, 1999b; Schwandt, 2018; Taut, 2007)
2. Identifying/challenging assumptions and values	Including discussions of evaluator or stakeholder positionality, cultural competency	(Buckley et al., 2015; Fierro et al., 2018; Patton, 2018; Preskill & Torres, 1999b; Schwandt, 2018; Vo et al., 2018; Wehipeihana & McKegg, 2018)
3. Systematic evidence/data collection and analysis	Including valuing evidence/data, analyzing what kinds of data is needed to answer a question, how it should be collected and analyzed	(Baker & Bruner, 2006; Buckley et al., 2015; Fierro et al., 2018; Patton, 2018; Preskill & Torres, 1999b; Vo et al., 2018)
4. Application of learning	For improvement or decision-making	(Baker & Bruner, 2006; Buckley et al., 2015; Fierro et al., 2018; Preskill & Torres, 1999b)
5. Judgement of value	Based on clearly understood criteria	(Patton, 2018; Vo et al., 2018)
6. Social learning	A social constructivist perspective	(Preskill & Zuckerman, 2003; Preskill & Torres, 1999b; Schwandt, 2018)
7. Contextual/situated learning	Here, this is the context of the organization or setting	(Fierro et al., 2018; Preskill & Torres, 1999b; Vo et al., 2018)
8. Integrated through all work	Through non-evaluation work, intentionally, and at all levels of the organization, as an every-day practice	(Baker & Bruner, 2006; Buckley et al., 2015; Patton, 2018; Preskill & Zuckerman, 2003; Preskill & Torres, 1999b; Schwandt, 2018; Taut, 2007)

Evaluative thinking (ET) is more than merely thinking done by evaluators. While ET can be learned from evaluators (Baker & Bruner, 2006; Buckley et al., 2015; Cousins et al., 2014; Fierro et al., 2018; Preskill & Torres, 1999b; Vo et al., 2018), particularly via process use (Patton, 1998; Preskill & Zuckerman, 2003) through collaborative and participatory evaluation approaches (Baker & Bruner, 2006; Buckley et al., 2015; Fierro et al., 2018; Preskill & Torres, 2000a; Preskill & Zuckerman, 2003; Weiss, 1998), it is certainly possible for evaluations to occur in organizations where evaluative thinking is uncommon and for evaluative thinking to exist in organizations that do few evaluations. In an organization where ET is uncommon, especially among decision-makers, the findings if evaluations may not be used effectively (Buckley et al., 2015; Patton, 2018; Vo et al., 2018). Buckley et al. (2015) call evaluative thinking, "the substrate that allows evaluation to grow and thrive" (p. 4).

For that substrate to effectively foster strong evaluation use, it must be spread beyond evaluators. One thing which distinguishes evaluative thinking in an organization from just doing good evaluations is its permeation through other elements of organizational functions. Preskill and Torres (1999b) and others emphasize the importance of ET becoming integrated into regular working processes throughout the organization, including and especially at the leadership and decision-making levels (Buckley et al., 2015; Duignan, 2003; Schwandt, 2018).

While Vo et al. (2018) and others suggest that evaluative thinking is something that evaluators should teach to non-evaluators, it may be dangerous to think of evaluators as the keepers of evaluative thinking. Preskill and Torres (1999b) remind that organizations are a political mix of many cultures. Evaluators have a culture of their own—with their own values, learning systems, language, etc.—and when evaluators work with organizations (or other staff within their own organization), it is akin to a cross cultural exchange. This exchange may be welcomed, or may feel colonial, with evaluators imposing the culture of evaluation over others (Patton, 2008). In their experience as evaluators working with indigenous communities, Wehipeihana and McKegg (2018) note knowledge systems are deeply cultural, with the culture or cultures in which people are embedded creating biases that may interfere with an evaluators ability to meet the needs of the evaluand. The authors would argue not only that western knowledge systems cannot adequately serve the needs of many indigenous communities, but that the elements common to indigenous thinking (including interconnectedness, the idea they knowledge belongs to the group, and that thinking, feeling, and doing are inseparable) could be beneficial to Western traditions of evaluation. This is where Vo et al.'s admonition to be cognizant of positionality is particularly appropriate. The presence and process of evaluation is not benign (Patton, 1998) and, like Schrödinger's cat, the very acts of observation and evaluation can change the subject, not always for the better.

What is an Evaluation Culture?

If, as has been suggested above, it is important for evaluative thinking to spread beyond educators to become an everyday aspect of the function of an organization, at what point does it become a characteristic of the institution? Buckley et al. (2015) suggest an organization which thinks evaluatively must be made of a critical mass of evaluative thinkers at all levels of the organization, even though staff at different levels might need to engage in evaluative thinking differently. Still, what constitutes a critical mass? Is a certain number or percentage of individuals all that is required, or are systematic changes to policies, infrastructure, and workplace culture necessary?

Buckley et al. consider evaluative thinking something done by individuals, but the definition adopted for this study follows the suggestion of Schwandt (2018) in recognizing the social nature of evaluative thinking. Schwandt reminds (a) the unit of analysis in an evaluation of any kind is almost always a group, (b) core evaluation activities like boundary decisions need to be answered collaboratively (i.e., it is not up to the evaluator's discretion), (c) especially in collaborative, participatory, and empowerment approaches, the evaluator is as much facilitator as judge. Schwandt characterizes the evaluation process of the latter approaches as a "communal sense-making process" (Schwandt, 2018, p. 132) with the focus being on answering the question: what should WE do? So, if evaluative thinking is already a social process that should be woven into the everyday practice of an organization, what is an evaluation culture?

"Every organization . . . has a culture of evaluation" (Murphy, 1999, p. 1). That culture might be supportive, fearful, or perhaps aspirational, but being a-cultural is not an option. The idea of culture is tied to a collection of shared values, practices, and norms in an organization that could be declared or implied. In the literature, the idea of an evaluation culture is closely linked to the concept of a learning organization. In some cases, the definitions are almost indistinguishable (Figure 1). Some definitions lean heavily in favor or elevating evaluation as a priority function of the organization (Murphy, 1999; Owen, 2003), while others support structures and practices that promote learning (Marsick & Watkins, 2003; Ortenbiad, 2002) and still others sound remarkably similar to definitions of evaluative thinking (Mayne, 2009; Preskill & Torres, 2000a).

In reviewing the literature, there are key elements or characteristics brought out in the discussion that may define an evaluation culture or support its development and sustainability. There were 31 papers reviewed with some discussion of either an evaluation culture or learning organization. There were 18 concepts mentioned more than twice as a key element of an evaluation culture/learning organization or an important contributor to the development or sustainability of an evaluation culture/learning organization. The results are summarized in Table 2. The most common concepts (mentioned in 17 of 31 references) were *using evaluation for improvement* and *leadership support*, followed by *resources available for evaluation* and *understands/accepts use of evaluation* (15 and 11 mentions, respectively). The idea of using evaluation for improvement reflects common concerns about the instrumental use of evaluation, but it also maps closely to definitions of evaluative thinking. The idea of learning for improvement or adaptation or evolution was most commonly associated with the definitions of learning organizations (see solid underlined terms in Figure 1), whereas the core function of evaluation is to render a judgement of merit/worth/significance (Scriven, 1991). It is only through *use* that evaluation findings may be applied for program improvement, adaptation, or even elimination. It would make sense then to think of an evaluation culture or learning organization--hereafter generally referred to collectively as an evaluation culture (EC)—as a setting where strong evaluation is conducted and likely to be used for program improvement and other decision-making.

Leadership support is a broad concept and in a different coding scheme could have also encompassed: resources available for evaluation, *policies and incentives*, and perhaps others. In this case, leadership or management support would wash out all other concerns.

Table 2

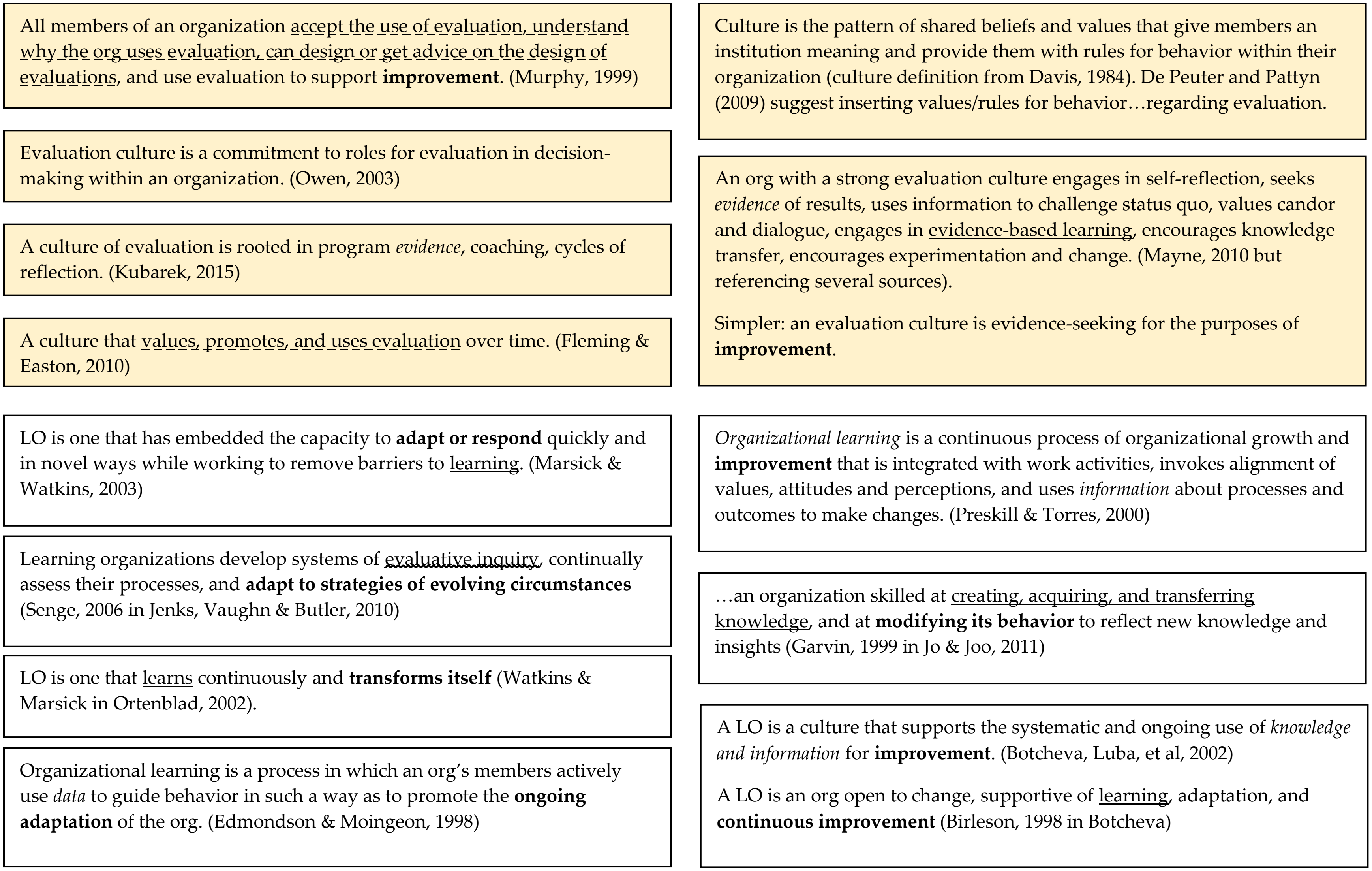
Characteristics Associated with an Evaluation Culture/Learning Organization

Characteristics	Coding	Count
1. Use evaluation for improvement	1	17
2. Leadership support	6	17
3. Resources available for evaluation (time, money, expertise)	9	15
4. Understand/supports organization's use of evaluation	23	11
5. Organization conducts evaluations	7	9
6. Transparency around purpose, communication	8	9
7. Staff regularly ask questions, participate in inquiry	18	8
8. Professional development in evaluation available	11	7
9. Staff or organization (internal) demands) for evaluation	5	6
10. Teams or communities of practice	16	6
11. Policies or incentive encourage evaluation use	15	6
12. Systems, systems thinking	17	5
13. Context awareness	19	5
14. Cycles of reflection	12	4
15. Openness to change, readiness to learn	20	4
16. Professional evaluation staff	10	3
17. Risk taking	13	3
18. Ownership of evaluation	14	3

Note. Literature reviewed: (Barnette, Wallis, & Barber Wallis, 2003; Botcheva, White, & Huffman, 2002; Carman & Fredericks, 2010; Coopey, 1995; Cousins et al., 2014; De Peuter & Pattyn, 2009; Duignan, 2003; Edmondson & Moingeon, 1998; Ewell, 2002; Fleming & Easton, 2010; Grudens-Schuck, 2003; Jenks, Vaughan, & Butler, 2010; Jo & Joo, 2011; Kubarek, 2015; Labin et al., 2012; Marsick & Watkins, 2003; Mayne, 2009; Murphy, 1999; Ortenbiad, 2002; Owen, 2003; Owen & Lambert, 1995; Preskill & Boyle, 2008; Preskill & Torres, 2000a; Preskill, Torres, & Martinez-Papponi, 1999; Sanders, 2002, 2003; Stufflebeam, 2002; Taut, 2007; Taylor-Powell & Boyd, 2008; Volkov & King, 2007; Williams & Hawkes, 2003)

Marsick and Watkins (2003) call leadership support the most important factor in facilitating a learning organization and "the [factor] most significantly related to perceived changes in financial performance" (p. 142). Leadership support has been tied variously to introducing evaluation or evaluative thinking to the organizations (Owen, 2003), taking ownership or responsibility for evaluation (Volkov & King, 2007), establishing an institutional value around learning and results management (Coopey, 1995; Marsick & Watkins, 2003; Mayne, 2009; Preskill & Boyle, 2008; Preskill & Torres, 1999b), creating structures that support learning capture and transfer (Preskill & Boyle, 2008; Preskill & Torres, 1999b), providing clear communication around the purposes, results, and decisions associated with evaluation (Marsick & Watkins, 2003; Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008), demonstrating accountability to evaluation results (Jo & Joo, 2011; Marsick & Watkins, 2003; Mayne, 2009; Preskill & Torres, 1999b; Taut, 2007; Taylor-Powell & Boyd, 2008), creating policies and incentives for participating in evaluative activities (Carman & Fredericks, 2010; Preskill & Boyle, 2008; Preskill & Torres, 1999b; Taut, 2007; Taylor-Powell & Boyd, 2008; Williams & Hawkes, 2003), and, of course, the availability of necessary resources (typically in the form of time, money, and expertise)(Fleming & Easton, 2010; Mayne, 2009; Murphy, 1999; Preskill & Boyle, 2008; Sanders, 2003; Stufflebeam, 2002; Volkov & King, 2007, and others). Note that the latter were also cited commonly among zoos and aquariums as reasons prohibiting the development of evaluation capacity (Clavijo et al., 2005; Khalil & Ardoin, 2011; Luebke & Grajal, 2011; Ogden & Heimlich, 2009; Roe et al., 2014). There are also negative associations with a *lack* of leadership support with some identifying this gap as a key factor in unsuccessful efforts to build evaluation culture in an organization (Preskill & Torres, 1999b; Sanders, 2003; Taut, 2007).

Below these top two concepts are two related constructs: *understand/supports the use of evaluation*, and *organization conducts evaluation*. As described, evaluative thinking is an ongoing, informal process with the purpose of improving the activities and therefore success of the organization. There are instances, however when a more formalized evaluation is necessary (required external reviews for grantors, program audits, etc.). Accepting this and considering the overwhelming call for leadership support in developing a culture of learning and improvement, then a definition can be formed for the purposes of this study where:



All members of an organization accept the use of evaluation, understand why the org uses evaluation, can design or get advice on the design of evaluations, and use evaluation to support **improvement**. (Murphy, 1999)

Evaluation culture is a commitment to roles for evaluation in decision-making within an organization. (Owen, 2003)

A culture of evaluation is rooted in program *evidence*, coaching, cycles of reflection. (Kubarek, 2015)

A culture that values, promotes, and uses evaluation over time. (Fleming & Easton, 2010)

Culture is the pattern of shared beliefs and values that give members an institution meaning and provide them with rules for behavior within their organization (culture definition from Davis, 1984). De Peuter and Pattyn (2009) suggest inserting values/rules for behavior…regarding evaluation.

An org with a strong evaluation culture engages in self-reflection, seeks *evidence* of results, uses information to challenge status quo, values candor and dialogue, engages in evidence-based learning, encourages knowledge transfer, encourages experimentation and change. (Mayne, 2010 but referencing several sources).

Simpler: an evaluation culture is evidence-seeking for the purposes of **improvement**.

LO is one that has embedded the capacity to **adapt or respond** quickly and in novel ways while working to remove barriers to learning. (Marsick & Watkins, 2003)

Learning organizations develop systems of evaluative inquiry, continually assess their processes, and **adapt to strategies of evolving circumstances** (Senge, 2006 in Jenks, Vaughn & Butler, 2010)

LO is one that learns continuously and **transforms itself** (Watkins & Marsick in Ortenblad, 2002).

Organizational learning is a process in which an org's members actively use *data* to guide behavior in such a way as to promote the **ongoing adaptation** of the org. (Edmondson & Moingeon, 1998)

Organizational learning is a continuous process of organizational growth and **improvement** that is integrated with work activities, invokes alignment of values, attitudes and perceptions, and uses *information* about processes and outcomes to make changes. (Preskill & Torres, 2000)

…an organization skilled at creating, acquiring, and transferring knowledge, and at **modifying its behavior** to reflect new knowledge and insights (Garvin, 1999 in Jo & Joo, 2011)

A LO is a culture that supports the systematic and ongoing use of *knowledge and information* for **improvement**. (Botcheva, Luba, et al, 2002)

A LO is an org open to change, supportive of learning, adaptation, and **continuous improvement** (Birleson, 1998 in Botcheva)

Figure 1. Definitions and Language from the Literature Related to Constructs of *Evaluation Culture* (shaded) and *Learning Organizations*. Common emphases (bold, italics, underlining) denote similarities.

> *An evaluation culture is one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions—with the mandate and support of organizational leadership.*

Thinking about an organization's evaluation culture is not about presence and absence. Remember, "[e]very organization . . . has a culture of evaluation" (Murphy, 1999, p. 1), but it may be weak or strong. Mayne (2008) offers an example of what a weak evaluation culture could look like:

> A weaker evaluative culture might:
>
> - gather information on results, but limit its use mainly to external reporting,
> - acknowledge the need to learn, but not provide the time or structured occasions to do so,
> - claim it is evidence-seeking, but discourages challenging and questioning the status quo, an/or
> - talk about the importance of achieving results, but value following the rules and frown on risk taking. (p. 1)

And so the question becomes: how does one measure the strength of an evaluation culture and, in the context of this study, how is it associated with the presence or absence of evaluators?

Internal Versus External Versus No Evaluators

Not working with evaluators is almost always a choice and rarely a consequence of circumstance. In any organization, resources are prioritized by leadership. Leadership may or may not decide to expend resources (time, money, expertise) on an evaluation function in favor of providing more programs and serving more constituents. However, while an organization may choose not to work with evaluators, it would be hard to imagine a context with zero evaluative activities, even if it were as simple as staff debriefing after program activities, personal reflections on success or failure, reports to supervisors, personnel evaluations, or program satisfaction surveys. These kinds of activities may be the early kernels of evaluative thinking, though Duffy (quoted in Volkov, 2011) would argue that these are *not* examples of internal evaluation, which the author would reserve for evaluative activities conducted by qualified and experienced staff. That said, there are certainly organizational circumstances where staff whose primary responsibilities may be management or programming, but who have training and experience in evaluation, conduct or facilitate internal evaluation activities.

At some point, an organization will choose or be required (by a funder or oversight agency) to conduct formal evaluation activities. These may be carried out by consultants or colleagues from peer institutions. External evaluators bring fresh eyes to a problem, unburdened by organizational history or intra-organizational relationships or dynamics. Of course, a good evaluator will learn these histories and dynamics, but they have less risk of allowing these elements to narrow their perspective. External evaluators represent a discrete investment of resources, which may even be built into the funding structure of a program grant and therefore represent a less daunting drain on existing organizational resources. Worthen, Sanders, and Fitzpatrick (2004) cite several advantages to working with external evaluators including the ability to target specific skills and experience that might be relevant to the evaluation at hand, and the potential for broader scale knowledge about how similar organizations operate (based on their earlier work). The authors also note that external evaluators will always have the enhanced credibility that comes with the assumption of third-party objectivity, but that financial relationships between the contractor and evaluator could compromise objectivity just as easily. It is fair to say that external evaluators trade fresh eyes for familiarity. It is difficult for an external evaluator to have as much knowledge of an organization's programming and personnel when compared to internal staff. It may also be difficult for external evaluators to build trust and rapport with staff, especially if skepticism of evaluation is common in the organization. Because external evaluators spend less time integrated into an organization and often have a finite engagement, it may be more difficult for them to contribute to evaluation capacity-building activities, including efforts to develop a strong evaluation culture. There are instances that blur the line between what might be considered internal and external. Long-term, ongoing relationships between evaluators and organizations can build the familiarity that might typically be lacking in an external consultant, but that familiarity could also hamper objectivity over time. Using the same evaluator with the same skillset also minimizes the benefit of matching the skills and experience of a consultant to the challenges of a particular program situation (Worthen et al., 2004).

Hiring internal evaluation staff is a demonstration of priority by leadership, one of the key aspects of leadership support associated with developing a strong evaluation culture. Advantages of internal evaluation staff according to Stufflebeam and Shinkfield (2007) include the ability to learn and understand staff and programs intimately, understanding the decision-making style of the organization, better tracking of results and data over time, and availability to provide training and technical assistance. The latter is a commonly cited resource lacking in programs with interest in using evaluation more regularly. Continuity leads to more consistent data collection, more evaluations conducted, and the ability to communicate results and data trends more consistently (Worthen et al., 2004). There are limitations to relying on internal evaluation staff as well. Internal staff may be less objective, or at least be seen that way by skeptics (Worthen et al., 2004), though periodic meta evaluations may bolster credibility (Stufflebeam & Shinkfield, 2007). If they are employed as regular staff, there are benefits and other associated employment costs (hiring, management) that are less associated with contract staff. There is a risk of staff abdicating their responsibility for evaluation or evaluative thinking ("not my job") and turnover in the position(s) can result in months-long interruptions of service. Worthen et al. also suggest that internal staff may struggle if not given enough authority or autonomy. Sonnichen (cited in Worthen, et al., 2004) suggests internal evaluators operate independently within the organization, with a high degree of autonomy and reporting to the top official. This, however, can be isolating. In one museum evaluator's experience, being forced to operate essentially outside of the organizational structure to maintain an appearance of objectivity led to them feeling isolated and ultimately less effective (K. Khalil, personal communication, April 22, 2020). They felt less able to take on some of the many roles that provide additional benefits to the organization. Picciotto (2013), however, argues that the independence of internal evaluators enhances the credibility and accountability of an organization, suggesting that independence is a core competency that needs to be developed by internal evaluators no matter where they fall within the organization. Picciotto suggests addressing the threat of structural isolation with strict protocols for professional interaction. Volkov (2011) identified eight interrelated roles commonly taken on by internal evaluators: (a) change agent, (b) educator about evaluation, (c) evaluation capacity building (ECB) practitioner,

(d) support for management decisions, (e) consultant, (f) researcher, (g) advocate, and (h) promoter of organizational learning. These broad responsibilities hint at the potential impact internal staff can have on the evaluation culture of an organization, though Worthen et al. caution against forcing internal staff to become *jacks-of-all-trades*, thus losing focus on their core responsibilities.

Many of the roles described by Volkov (e.g., educator, advocate, ECB practitioner, promoter of learning) could contribute to building the capacity for evaluative thinking and a stronger evaluation culture. In fact, Volkov concludes by calling for internal staff to work to develop an *evaluation meme* in their institution. Beere (2005) describes evaluation capacity building as the responsibility of an internal evaluator (or staff)—particularly increasing demand for evaluation within the organization—because of their position and credibility with staff. García-Iriarte, Suarez-Balcazar, Taylor-Ritzler, and Luna (2011) describe an opportunity for internal evaluators to serve as a *catalyst for change*. If internal staff have sufficient authority (or position within the organization), are given or can develop influence among staff and leadership about the role of evaluation within the institution, then they can be effective change agents by using a combination of evaluation participation (process use) and evaluation education. The success of this approach hinges on the political acumen of the individual, the support from leadership, and the size and complexity of the organization. Progress is also vulnerable to staff turnover.

Clearly, internal versus external evaluators is not an either/or proposition. Taut (2007) and Volkov (2011) cite multiple studies that support the enhanced benefits of balancing or supplementing internal and external evaluation, providing check and balances and maximizing benefits of each while minimizing challenges.

Evaluation is growing in influence within zoos and aquariums as a tool to improve their conservation mission performance. This trend is fueled by professional accreditation standards and industry norming. Institutions face choices in how they chose to engage in evaluation activities, including working with professional evaluators on a contract basis, hiring internal evaluation staff, going it alone, or some combination of the three. However, evaluation for evaluation's sake may not have the same institutional impact as the strategic development of evaluative thinking in staff in a culture of evaluation that promotes reflective practice and data-driven, context-appropriate decision-making. It is clear that evaluators, and especially internal evaluators, have the potential for promoting evaluative thinking and the development of a learning organization, but what is not clear from the literature is how working with evaluators is related to the state of an organization's evaluation culture. This is the question this study addresses, which has implications in the realm of both theory (on the development of evaluative thinking and evaluation culture) and practice (how institutions can best utilize evaluators towards the improvement of mission performance).

Chapter Three: Methods

This section describes the two-phase, mixed-methods design used in this study, including the characteristics of the study respondents. Data collection included use of a survey instrument followed by case study interviews with a sub-set of participants. The independent and dependent variables are described, and the data analysis plan is presented for each study phase and the synthesis.

Respondents

The Association of Zoos and Aquariums (AZA) is a professional organization of member institutions "dedicated to the advancement of zoos and aquariums in the areas of conservation, education, science, and recreation" (Association of Zoos & Aquariums, n.d., paragraph 1). Members are required to pass a strict accreditation process that includes standards for education programming and program evaluation (an excerpt from the accreditation standards is provided in Table 3). Standards address a broad array of institutional parameters with an emphasis on safety and animal welfare. It is possible to be accredited without meeting every standard and so the conduct of evaluation activities throughout member institutions is variable. There were 233 accredited institutions at the onset of this study, including 215 in the United States (Association of Zoos & Aquariums, 2018b). Participants for the present study were drawn from the 215 U.S. institutions.

Table 3

Sample Accreditation Standards (Association of Zoos & Aquariums, 2019)

Standard	Description
4.3.1.	Classes, programs, animal talks, interpretive programs and other education programs should be evaluated on a regular basis for effectiveness and content. Programs should be updated with current scientific information, with an educational/conservation message as an integral component. These evaluations should assess more than participant satisfaction, looking also at program impact (ideally including impact on conservation-related knowledge, attitudes/affect, and behavior). *Results from evaluations should be used to improve the existing programs and to create new programs.*
4.3.2.	The institution should have a thorough understanding of the needs of its audiences and as such provide programs to meet these needs.

Note. Italics denote emphasis added by author

The following variables were tracked to describe the characteristics of zoos and aquariums: governance, institutional budget, and annual attendance. These are summarized in Table 4. AZA categorizes institutions as for-profit, non-profit, or public (government/municipal) according to their operating authority. Some institutions, like the Seattle Aquarium, may be owned by a governmental entity (in this case, the City of Seattle), but are operated by a non-profit (the Seattle Aquarium Society). This example was included with non-profits. Institutional size was determined in two ways, budget and attendance. Institutional budgets are broken into four categories annually for benchmark reporting by AZA: small (< $2 million operating budget), medium ($2 million-$6.9 million), medium-large ($7 million-$26 million), and large (> $26 million). Annual attendance is not necessarily linked to operating budget because there are numerous free zoos and aquariums that generate high attendance numbers with smaller operating budgets. Dividing AZA institutions into quartiles by 2017 annual attendance provided four categories that mirror the annual budget categories: small (< 100,000 annual visits), medium (100,000-299,999), medium-large (300,000-600,000), large (> 600,000).

Institutional data on budget and attendance came from AZA's 2018 benchmarking reports (Association of Zoos & Aquariums, 2018a). Information in these reports is anonymous (institutions are described as "Institution ##"). The type of institution (zoo, aquarium, other) is readily available, but was not specifically included in data collection as it was unlikely to have relevance to the research questions. There were several very large institutions among the member associations that had attendance numbers twice the size of the average institution in the largest attendance category and as much as 10 times the number of staff. There were less than 10 institutions in this outlier group operated by two corporate entities. These ultra-large organizations were excluded from the sample.

Table 4

Respondent Variables

Variable	Category	Description
Governance	For-profit	Zoo/aquarium operated as a for-profit business
	Non-profit	Zoo/Aquarium operated as a 501c3 or similar non-profit tax status
	Public	Zoo/Aquarium both owned and operated by government or municipality
Budget	Small	Less than $2 million annual operating budget
	Medium	$2-6.9 million annual operating budget
	Medium-large	$7-26 million annual operating budget
	Large	Greater than $26 million annual operating budget
Attendance	Small	Fewer than 100,000 annual visits
	Medium	100,000-299,999 annual visits
	Medium-large	300,000-600,000 annual visits
	Large	Greater than 600,000 annual visits

Note. Categories adapted from AZA benchmarking reports (Association of Zoos & Aquariums, 2018a)

The unit of analysis was the education/engagement department or working group within the institution. These are the working groups where evaluation activities are most likely to be located and the working groups where an effect was most likely to be seen. While it is possible the education and exhibit or interpretive functions may be separated in larger organizations, the focus in this study centered on the education/engagement departments with additional context explored in phase two when institutions were included where these functions were separated. Work/working group and department are used interchangeably throughout. Only a single response was recorded for each institution.

Study Design

The study was conducted in two phases. First, a survey of education/engagement directors was conducted. Second, nine case-study institutions were identified for follow-up interviews. In this study, *director* refers to the senior department manager. Depending on the organization and their hierarchy, that role may be referred to as a director, curator, or vice-president. The education/engagement department was defined as the department where most instructor-led programming is based. Some interpretive functions may be separated into guest experience departments at some institutions, but to make a stronger comparison, this study focused on the education/engagement departments. In the final sample, some institutions made their own judgements on which department to put forward and so the sample includes some interpretive or public programming departments.

Being an education director myself comes with both advantages and challenges. I have connections throughout the industry and know many of the directors who might have elected to participate in the study (or they may have recognized me from conference presentations, committee work, articles, or other AZA activities). This may have led to better response rates, but may have also exacerbated social desirability bias or satisficing. It may have also led to more candid responses in case-study interviews. My perspective and experience also meant that I had a rapport with interview participants, and we shared a clear understanding of the work. This should also make any recommendations that arise from the study more practical and realistic. On the other hand, my personal experience as an education director could limit my perspective to what has worked (or not) for me in the past. This tension was something I carried and reflected on frequently throughout the design, conduct, and analysis of the study and its findings. One thing that would have been valuable would have been to keep a reflexivity journal. Hobson (2001) recommends keeping a journal during action research to bring awareness to the researcher's experience, activities, and perspective. While this is study is not action research, a reflexive journal could help any researcher become aware of patterns of thought or unconscious bias that may creep into the design or analysis of the work.

The risk for participation in the study was minimal. For the education directors, participation was entirely voluntary. The identity of the education directors is known to the investigator, but their survey and interview responses have been anonymized by site (e.g., Institution A) for the purposes of reporting. All information pertaining to the identity of study participants is held on a secured University cloud server. Informed consent was sought for participating education directors (at the beginning of the survey, see Appendix A), which included both phases one and two. The study protocol was approved by the University of South Florida's Institutional Review Board prior to the commencement of any participant outreach or data collection (see Appendix B).

Phase one: Survey. The data collection survey instrument was based on the Readiness for Organizational Learning and Evaluation instrument (ROLE) (Preskill & Torres, 2000b). There are six dimensions of the ROLE instrument: culture, leadership, systems and structures, communication, teams, and evaluation with 8 to 27 items per dimension (some with sub-categories within dimensions). Most items are scored on a 5-point Likert scale (strongly disagree to strongly agree) with three dichotomous items in the *teams* subscale (yes/no). Items were developed based on a literature review of organizational learning and evaluation readiness (including existing assessment tools) and a series of interviews with four organizations interested in enhancing their learning and evaluation processes and systems. Reliability across the 78 items was assessed by Preskill et al. (1999) with Cronbach's alpha and found to be strong (α = .97). Internal consistency for each of the six dimensions was also strong with alphas ranging from .83 to .94 (Preskill et al., 1999). See Appendices C for the original instrument and D for the modified instrument.

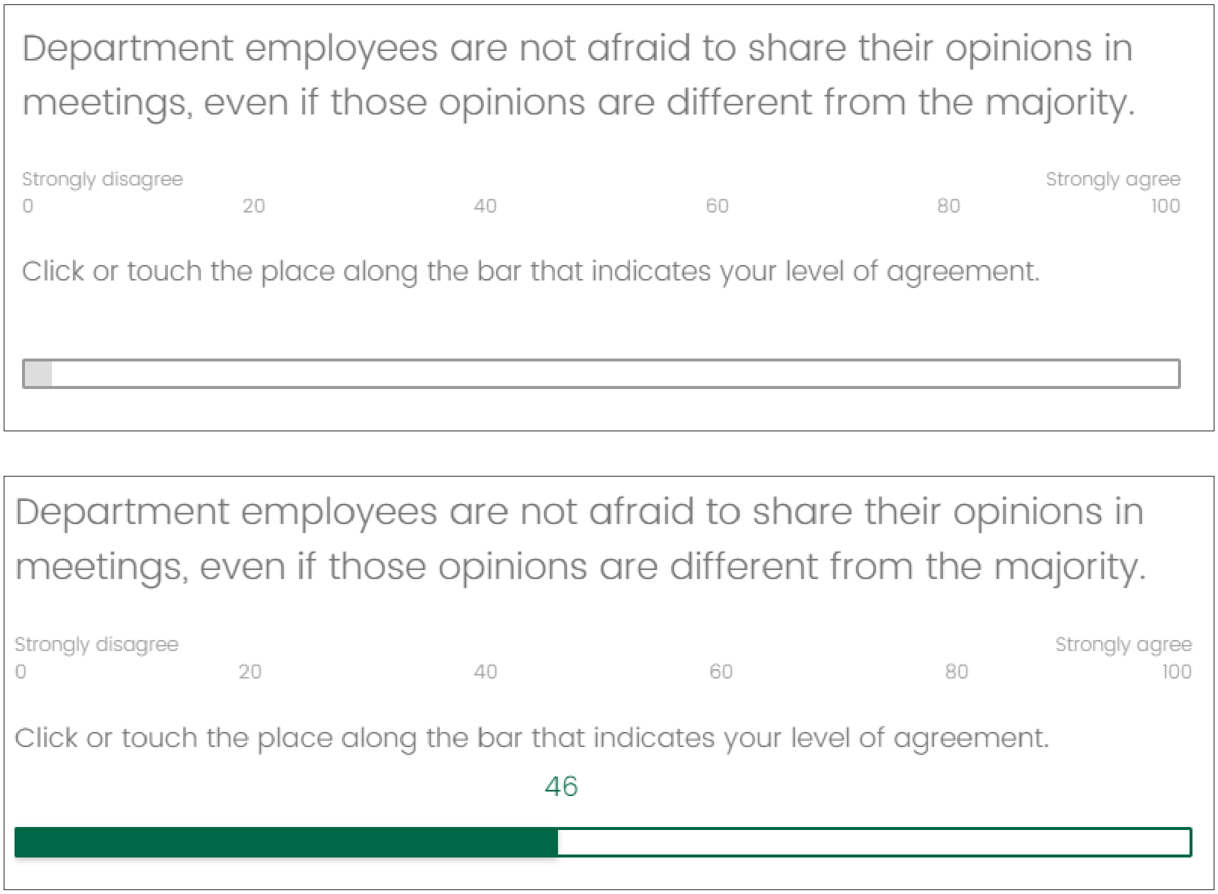
Survey modifications and review. Permission was sought from Preskill and Torres (see Appendix E) to modify and use the ROLE instrument, including publication as part of this dissertation. ROLE was modified initially to collect study-relevant demographic data, to obtain information about interactions with professional evaluators, and to inquire about openness to participate in the second phase of the study. The survey content and associated consent materials were uploaded to Qualtrics for distribution. Five professional colleagues with experience as professional evaluators in zoo and aquarium settings and/or as social science researchers agreed to serve as reviewers for the survey instrument (see Appendix F). A reviewers' version was sent to these colleagues with some additional framing information and a request to review the survey design, as well as face and content validity. The reviewers' version included questions at the end of the survey to specifically address these questions (included in Appendix F). Three of the five reviewers found the survey too long. Other design feedback included the addition of back buttons, clarification of language, some question design concerns (use of *and* in some questions). Reviewers generally agreed that the survey adequately addressed institutional demographics and work with professional evaluators (with some suggested changes). Reviewers were split on whether the survey adequately addressed the study concepts of evaluation culture and evaluative thinking. Concerns were expressed about whether the questions went into enough depth or had enough nuance and whether the survey design (including headers, inclusions of definitions, question wording) would result in respondents satisficing. The following changes were made to address reviewer concerns:

- Survey items were reviewed for alignment with study constructs and repetition. There were 25 items removed to reduce the primary question block to 50 items.
- Definitions of constructs and section headers were removed, and questions were condensed into a single 50-question, randomized block to reduce cuing.
- Language was modified throughout to clarify context (specifically, that respondents should be answering in the context of their department rather than the institution as a whole).
- Page breaks were added to minimize scrolling and a progress bar and back buttons were added to give respondents more control over their experience.
- The survey response format was changed from a 5-point, Likert-style scale to a 0-100 scale to increase item variance and reliability. All questions were placed on this scale.

Item design included bars for responses (see Figure 2). Respondents clicked or dragged along a solid bar to indicate their level of agreement with each statement.

Respondents could easily modify or change their response by clicking another location along the bar or dragging the endpoint to the desired location. In a review of the pertinent literature, Chyung, Swanson, Roberts, and Hankinson (2018) not several advantages of using continuous rating scales over versus discrete rating scales (e.g., Likert-type scales using radio buttons). They allow for precise ratings; they improve reliability by increasing item variance, and they tend to general more normally distributed data. They may also be more engaging for respondents, thereby reducing survey fatigue. Continuous scales my take the form of sliders or visual analog scales (VAS) like the bars used in this study. Multiple studies have questioned the use of sliders noting they take longer to complete (Funke, Reips, & Thomas, 2011; Roster, Lucianetti, & Albaum, 2015), lead to more incomplete surveys (Funke, 2016), and can be influenced by the starting place of the slider (at the beginning, middle, or end)(Buskirk, 2015). VAS scales do not have the challenge of slider start points and have similar performance to radio buttons (Toepoel & Funke, 2018) in regards to survey completion and response times. Matejka, Glueck, Grossman, and Fitzmaurice (2016) further suggest reducing bias in VAS scales by removing tick marks to prevent clumping around the locations of the marks.

A summary of reviewer feedback and specific changes made to the instrument can be found in Appendix F. The updated version of the instrument was shared with reviewers with a summary of changes. The final survey (Appendix D) was distributed to education department leadership staff at the Seattle Aquarium to review the surveys for clarity and ease of use. Suggested changes at this stage were largely transcription and typographical errors.


Department employees are not afraid to share their opinions in meetings, even if those opinions are different from the majority.

Click or touch the place along the bar that indicates your level of agreement.

Department employees are not afraid to share their opinions in meetings, even if those opinions are different from the majority.

Click or touch the place along the bar that indicates your level of agreement.

46

Figure 2. Item Design. Respondents indicated their level agreement by clicking or dragging along a solid bar. The upper image shows the item with no response. The lower image shows a resonse of 46 out of 100.

Independent variable. The independent variable in this study addressed work with professional evaluators. Work with professional evaluators can take many forms. Categories of work with internal and external evaluators are summarized in Table 5. Internal evaluation resources may include one or more program evaluators, social science researchers (staff conducting research on learning or behavior change), or audience researchers (staff who work directly with visitor audiences evaluating visitor experience rather than program impact). Internal evaluation resources may also include knowledgeable/experienced staff with college-level education or professional training in evaluation (not necessarily a degree) and/or those who have worked as evaluators in a previous position. External evaluation resources include contract or consultant evaluators hired to evaluate a specific program or project, external audience researchers (including firms that do public opinion research around the visitor experience or who may operate a kiosk or deploy staff to answer visitor experience questions for an institution) and university/research partnerships (including working with researchers or students learning or conducting evaluation or social science research).

Table 5

Categories of Work with Professional Evaluators

Internal Evaluation Resources	External Evaluation Resources
Program evaluators(s)	Contract/consultant (frequency)
Social science researchers	Audience researchers (e.g., Impacts, Morey)
Audience researchers	University/research partnerships
Knowledgeable internal staff	Other

Data related to interactions with evaluators were initially collected through the ROLE instrument in phase one. These data were clarified, and more detail was provided for selected case study institutions during the director interviews in phase two (including information gleaned from supplementary documentation).

Dependent variable. The dependent variable in this study was evaluation culture, defined in this context as: one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions . . . with the mandate and support of organizational leadership. The component constructs are: (a) instrumental use of evaluation, (b) evaluative thinking, and (c) leadership support.

First, are formal evaluations conducted to improve programs? To answer this question, this study used the modified ROLE instrument in the initial phase of the study and director interviews in the second phase of the study. The Evaluation section of the ROLE instrument (questions 43-50) addressed this construct directly. During the review process, two questions from the original instrument were removed to address concerns about length. Two questions (questions 49 and 50) were later added to balance the constructs tested and to address questions about the purposes of evaluation efforts and the consistency of data collection efforts more directly. See Table 6 for an enumeration of the questions addressing the Evaluation dimension.

Table 6

Questions Addressing the Evaluation Dimension in the ROLE Instrument

Question	Question Text
43	The integration of evaluation activities into our department's work has enhanced (or would enhance) the quality of decision-making.
44	Managers and supervisors in the department like (or would like) staff to evaluate their efforts.
45	Evaluation helps (or would help) the department provide better programs, processes, products and/or services.
46	There would be support among department employees if we tried to do more (or any) evaluation work.
47	Doing (more) evaluation would make it easier to convince department and organizational leadership of needed changes.
48	There are evaluation processes in place that enable department employees to review how well changes we make are working.
49	When the department engages in evaluation activities, the goal is to improve programs.
50	Data are routinely collected during department activities to inform evaluation efforts.

Note. All questions on a 0-100 response scale from *strongly disagree* to *strongly agree*.

Second, is there evidence of evaluative thinking among members of the workgroup? In this study, evaluative thinking is defined as: *a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making*. Component constructs within this definition include social learning, reflective practice, investigation of assumptions, systematic data collection, and context-appropriate decision-making. Each of these constructs are assessed by the ROLE instrument in phase one of the study and in the case-study interviews during phase two. Social constructivism and team learning were core concepts that contributed to Preskill and Torres's development of their model of organizational learning (on which the ROLE instrument is based) (Preskill & Torres, 1999a). Social learning was addressed in questions covering Teams (questions 38-42), as well as in the Systems and Structures and Culture sections (Tables 7, 8, and 9, respectively). Questions addressing assumptions are found in the section on Culture (e.g., "In meetings employees are encouraged to discuss the values and beliefs that underlie their opinions."). These ideas were drawn out and elaborated on during the case study interviews, particularly regarding positionality and cultural competence/relevancy/ bias. Questions on data collection and use in decision-making are found in sections on Culture, Leadership (Table 10), Communications (Table 11), and Evaluation. Leadership support includes the perception of appropriate resources (time, money, and expertise) and is fully operationalized by the ROLE instrument in phase one through sections specifically on Leadership (questions 21-28) and Systems and Structures. Evidence for leadership support was also explored during the case study interviews.

Table 7

Questions Addressing the Teams Dimension in the ROLE Instrument

Question	Question Text
38	Our department currently operates via (or is transitioning towards) a team-based structure where work projects are intentionally assigned to work groups rather than individuals with shared accountability and leadership.
39	Department employees are provided adequate training on how to work as a team member.
40	Team meetings in the department address both team processes and work content.
41	Team meetings in the department strive to include everyone's opinion.
42	Teams and work groups in the department are encouraged to learn from each other and to share their learning with others.

Note. All questions on a 0-100 response scale from *strongly disagree* to *strongly agree*.

Table 8

Questions Addressing the Systems and Structures Dimension in the ROLE Instrument

Question	Question Text
29	There is little bureaucratic red tape when trying to do something new or different in the department.
30	There are few boundaries between department units or working groups that keep employees from working together.
31	Department employees are recognized or rewarded for learning new knowledge and skills.
32	Department employees are recognized or rewarded for helping solve organizational problems.
33	The current reward or appraisal system in the department recognizes, in some way, team learning and performance.
34	Asking questions and raising issues about work with department leaders is encouraged.
35	Department employees are recognized or rewarded for experimenting with new ideas.

Note. All questions on a 0-100 response scale from *strongly disagree* to *strongly agree*.

Table 9

Questions Addressing the Culture Dimension in the ROLE Instrument

Question	Question Text
1	Department employees respect each other's perspectives and opinions.
2	Department employees ask each other for information about work issues and/or activities.
3	Department employees continuously look for ways to improve processes, products and/or services.
4	Department employees are provided opportunities to think about and reflect on their work.
5	Department employees often stop to talk with each other about the pressing work issues we're facing.
6	When trying to solve problems, department employees use a process of working through the problem before identifying solutions.
7	Department employees operate from a spirit of cooperation, rather than competition.
8	Department employees tend to work collaboratively with each other.
9	Mistakes made by department employees are viewed as opportunities for learning.
10	Department employees continuously ask themselves how they're doing, what they can do better, and what is working.
11	Department employees are confident that mistakes or failures will not affect them negatively.
12	Managers and supervisors in the department view individuals' capacity to learn as among the organization's greatest resources.
13	Department employees use data/information to inform their decision-making.
14	Asking questions and raising issues about work with department leaders is encouraged.
15	Department employees are not afraid to share their opinions in meetings, even if those opinions are different from the majority.
16	Department employees feel safe explaining to others why they think or feel the way they do about an issue.
17	Department employees are encouraged to take the lead in initiating change or in trying to do something different.
18	Managers and supervisors in the department make decisions after considering the input of those affected.
19	In meetings, department employees are encouraged to discuss the values and beliefs that underlie their opinions.
20	Department employees are encouraged to offer dissenting opinions and alternative viewpoints.

Note. All questions on a 0-100 response scale from *strongly disagree* to *strongly agree*.

Table 10

Questions Addressing the Leadership Dimension in the ROLE Instrument

Question*	Question Text
21	Managers and supervisors in the department take on the role of coaching, mentoring and facilitating employees' learning.
22	Managers and supervisors in the department help employees understand the value of experimentation and the learning that can result from such endeavors.
23	Managers and supervisors in the department are open to negative feedback from employees.
24	Managers and supervisors in the department model the importance of learning through their own efforts to learn.
25	Managers and supervisors in the department believe that success depends upon learning from daily practices.
26	Managers and supervisors in the department support the sharing of knowledge and skills among employees.
27	Managers and supervisors in the department provide the necessary time and support for systemic, long-term change.
28	Managers and supervisors in the department use data/information to inform their decision-making.

Note. All questions on a 0-100 response scale from *strongly disagree* to *strongly agree*.

Table 11

Questions Addressing the Communications Dimension in the ROLE Instrument

Question*	Question Text
36	Information is gathered from clients, customers, suppliers or other stakeholders to gauge how well we're doing.
37	There are adequate records of past change efforts and what happened as a result.

Note. All questions on a 0-100 response scale from *strongly disagree* to *strongly agree*.

Survey distribution. The survey was distributed electronically via Qualtrics directly to the education contact at each AZA institution on July 29, 2020. This mailing list was provided by AZA in support of the study (see letter of support in Appendix E). A number of strategies were employed to maximize response. Fan and Yan (2010) identified four stages of web survey development and distribution that can influence response rate: development, delivery, completion, and return. Several factors identified in the development stage were addressed in the design of the survey, including sponsorship, content salience, and length. By providing the electronic mailing lists, AZA provided an implied endorsement. Using my institutional affiliation further lent credibility to the effort for my AZA peers (Fan & Yan, 2010; Sheehan, 2001). The content was particularly relevant (evaluation and evaluation culture), because these have been topics of interest among the community, as evidenced by the prevalence of conference sessions and workshops at recent AZA mid-year and annual conferences. Length was addressed by reducing the number of items and editing the text to remove study definitions and section headers. In the delivery phase, several studies recommend the use of pre-contact communications to improve response rate by alerting the respondents of the arrival of the survey . . . which also helps circumvent spam filters (Fan & Yan, 2010; Kaplowitz, Lupi, Couper, & Thorp, 2012; Sheehan, 2001). Personalization is another common recommendation (Fan & Yan, 2010; Sauermann & Roach, 2013). Qualtrics allows for creating personalization in its email and communication platforms. Correspondence used the first names and institutions of contacts in their e-mails. However, this ease of customization is well-known and therefore may minimize this value (Fan & Yan, 2010; Muñoz-Leiva, Sánchez-Fernández, Montoro-Ríos, & Ibáñez-Zapata, 2010; Porter & Whitcomb, 2016). Surveys, especially institutional surveys like this one, may also commonly be forwarded, which also reduces the value of personalization (Monroe & Adams, 2012). The invitation letter, distributed to 206 education contacts, used language suggesting participants were *part of a select group* (Porter & Whitcomb, 2016), was clear and honest about the time required for completion (Fan & Yan, 2010), placed the link at the bottom (Kaplowitz et al., 2012), and asked for help in the subject line (Trouteaud, 2004). All survey correspondence can be reviewed in Appendix G. Several surveys were returned as undeliverable. For these institutions, AZA provided access to an additional mailing list consisting of institutional directors (CEOs, presidents, etc.). A supplementary invitation was sent to 11 institutions. Several respondents indicated that they did not receive the survey link after receiving the pre-invitation letter. These instances were handled individually through a variety of means to ensure they could complete the survey (e.g., sending to personal email addresses, sending a direct link). Reminder emails were sent one and two weeks after the initial distribution on August 3, 2020 and August 12, 2020. The reminder on August 12 extended the two-week deadline to maximize return. A targeted reminder was sent to 25 respondents who had started the survey, but not completed it. The survey closed on August 19, 2020. This completed data collection for phase one of the study.

Phase two: Case-study interviews. The second phase of the study involved an interview with the education/engagement director of nine case study organizations (*N* = 9). To identify institutions for inclusion in the case study, an overall evaluation culture score was calculated from the responses to the modified ROLE survey instrument (Appendix D). This was accomplished by creating a mean score for each dimension then averaging the dimensional means for an *overall evaluation culture score*. The maximum possible score directors could give their organizations is 100 (indicating a rating of 100, or strongly agree, on a 0-100 scale on every item in each dimension); the lowest would be 0 (indicating a rating of 0, or strongly disagree, on all items). The median score would be 50 (an average rating of 50 on each scale, indicating neither agreement nor disagreement). Scores for each organization were separated into three categories (strong, moderate, and weak) based on the responses of the education director. The tiers were constructed by dividing the 100 scores into thirds (33 scores in the top/strong tier, 34 scores in the middle/moderate tier, and

33 scores in the bottom/weak tier). Other approaches for creating these categories were considered. Using scores above and below one standard deviation from the mean resulted in too few cases in the strong and weak categories (*n* = 17) for identifying follow-up cases. Using scores one standard deviation around the median (overall evaluation score of 50) created an empty weak category (*n* = 0) and an over-represented strong category (*n* = 97). Identifying natural breaks in the data resulted in too few cases in the weak category (*n* = 17) for identifying follow-up cases.

From these categorized responses, three case studies were identified in each tier (strong, moderate, and weak) that had the most explanatory potential. Details for each institution are summarized in Table 12. In the strong tier (the upper third of overall evaluation culture scores), the following institutions were selected:

- Institution E: A large, well-resourced institution with a large internal evaluation staff that scored itself highly in all categories (subscales, overall score, and the emergent construct, psychological safety). Overall evaluation culture score: 89.59 (out of 100).
- Institution B: A small institution with no internal evaluation staff, limited exposure to external evaluators that also scored itself highly in all categories. Overall evaluation culture score: 84.72.
- Institution I: A medium-large institution with high scores but no indicated work with internal or external evaluators. Overall evaluation culture score: 82.17.

The second tier (middle third of scores) included these institutions:

- Institution F: A large institution with some internal evaluation capacity indicated. Overall evaluation culture score: 79.59.
- Institution D: A medium-large institution scoring right at the mean that indicated no internal evaluation staff but frequent work with external evaluators. Overall evaluation culture score: 76.79.
- Institution H: A small institution indicating infrequent work with external evaluators and no internal evaluation staff. Overall evaluation culture score: 74.76.

Finally, in the third tier (lower third of evaluation culture scores):

- Institution C: A large, well-resourced institution that indicated both internal evaluation staff and work with external evaluators, but that score itself poorly in many categories (relative to other institutions). Overall evaluation culture score: 69.69.
- Institution G: A small institution with limited work with external evaluators and no internal evaluation staff. Overall evaluation culture score: 60.49.
- Institution A: A medium-large, public institution that works with external evaluators but was among the lowest scoring institutions. Overall evaluation culture score: 55.58.

Institution Y and Institution Z were initially identified for the top and middle tiers, respectively. Institution Y was not responsive to requests to participate in phase two and was replaced with a comparable institution (Institution I). Institution Z had a complicated and atypical relationship with a partner non-profit for programming and felt uncomfortable participating. They were replaced by an institution of similar size and with a similar relationship with evaluators (Institution H).

Table 12

Institutions Chosen for Case Study Interviews

			Overall	Work with Professional Evaluators		
Institution	Governance	Size	Evaluation Score	Internal	External	Trained
Tier 1 (strong)						
Institution E	Public	Large	89.59	Yes	Yes	Yes
Institution B	Non-profit	Small	84.72	No	Yes	No
Institution I	Non-profit	Med-large	82.17	No	No	Unsure
Tier 2 (moderate)						
Institution F	Non-profit	Large	79.59	Yes	Yes	Yes
Institution D	Non-profit	Med-large	76.79	No	Yes	No
Institution H	Public	Small	74.76	No	Yes	No
Tier 3 (weak)						
Institution C	Non-profit	Large	69.69	Yes	Yes	No
Institution G	Public	Small	60.49	No	Yes	No
Institution A	Public	Med-large	55.58	No	Yes	No

Note. Governance options include for-profit, non-profit, public (government, municipal). Institutional size options include small, medium, medium-large, large (categorized by annual budget). Overall evaluation score is out of a possible 100 (*M* = 74.79). Work with professional evaluators indicates presence/absence of internal evaluators, external evaluators, and internal (non-evaluator) staff with training comparable to a professional evaluator.

In preparation for the interview, each case study site was requested to provide their most recent accreditation materials related to education and evaluation, as well as any internal documents relevant to the topic. They include: the institutional education plan (AZA Accreditation Standard EI-2), audience needs assessment (EI-7), evaluation plan (EI-9), interpretive plan (EI-10), and Mission and Messaging Plan (internal document). Of particular interest was the evaluation plan (EI-9). The actual documents from participating organization are not included to protect the anonymity of the participants. Documents were provided by six of the nine case study institutions, three of which provided only their EI-9. These were reviewed to provide context for the director interviews. An interview protocol was developed to guide each interview that was based on the survey content (work with professional evaluators, definitions and examples of evaluation culture and evaluative thinking) and included questions related to an emergent concept from the phase one data analysis around the relationship of psychological safety to evaluation culture and evaluative thinking. The interview protocol was sent for review to three of the five evaluation professionals who reviewed the survey. All three were working social science researchers with experience in qualitative methods in the zoo/aquarium industry context. Their previous review of the survey instrument also made them familiar with the study, its goals, and methods. Reviewers suggested mostly moderate changes to language and approach to encourage richer responses and improve clarity. The final protocol and a summary of changes can be found in Appendix H.

The department director of each institution was contacted to secure their participation in the second phase of the study, including requesting the submission of relevant documents. A summary of their survey responses compared to the sample was provided to help them prepare for the interview (see Appendix I). An appointment was set for a one-hour videoconference interview. Documents were provided by Institutions C, E, H, and I prior to our conversations. Institutions F and G provided accreditation documents related to their evaluation plans (EI-9) after the interview, but they did not provide context or information useful beyond that collected in the survey and/or interview. Institutions A and D felt their EI-9s were too dated to be relevant. Institution B did not provide documents. Interviews were conducted over the course of two weeks. Each interview was recorded (with permission) via the videoconference software platform. The platform provided an auto transcript of the recording. Each transcript was reviewed and required extensive editing and corrections from the automated document. Interview transcripts were then organized and sorted following the protocol developed in Ose (2016). Ose's process uses spreadsheets to sort and organize the interview content for coding and analysis. Transcripts were formatted so that interviewer questions (marked I:) alternate with respondent comments (marked R:). The text was then transferred to a spreadsheet with interviewer and respondent comments on alternate rows; interviewer questions and comments were formatted in bold text for ease of review. The nine interviews were separated into nine worksheets within the spread. A 10th worksheet contained a list of interview subjects and their designations (Institution A, B, etc.). Interviews were then coded using an emergent thematic coding approach (Gibbs, 2007). Each question and response were coded together. When a response contained content appropriate for multiple codes, the question/response pair was duplicated in the spreadsheet. A code list was kept in a separate worksheet within the spreadsheet. As each interview was completed, previous interviews were reviewed to add or re-code responses according to new emergent codes. After the final interview, this amounted to a full review of coded statements for accuracy and consistency. Upon completion, 102 unique codes were identified with at least one question/response associated (Table 13). After coding, interviews were combined into a single sheet with the institution identified and code associated with each row/response. A series of formatting steps resulted in the creation of a word processing document with the responses sorted by codes. This document allowed for codes to be sorted and organized easily through the outline function of the word processing software. The codes organized by emergent themes can be found in Table 18 in Chapter Four. Interview responses are not included in this document as they contain information that would allow for identification of interview participants, even after names and other identifying elements are removed.

Table 13

Interview Codes

Code Description	Code
Defining terms/study clarifications	1
Position/department description/organizational circumstance	2
Professional evaluator definition	3
Ways respondent/department/org has worked with professional evaluators	4
Work with contract evaluators	5
Contract evaluators primarily for grants/special projects	6
Frequency of work with specific evaluators	7
Work with external audience researchers (including market research)	8
Director background/training in evaluation	9
Director training primarily informal/learn on the job	10
How has working with evaluators changed views in evaluation?	11
Haven't worked with evaluators enough for there to be influence*	12
Respondent ideas about evaluation culture	13
Study definition of evaluation culture	14
Evaluation culture/value/use can live in pockets w/in department/org	15
Respondent judgement of department/org evaluation culture	16
Department/staff value evaluation	17
Struggle to make time/find resources	18
Lack skills/facility to incorporate evaluation into processes	19
How does evaluation culture differ in dept vs. rest of org?	20
Org leadership value evaluation	21
Disagreement/differences in metrics	22
Review of survey scores	23
Communication scale	24
Systems and structures scale	25
Would staff score the survey differently?	26
Survey scores by staff would likely be very similar	27
Survey scores by staff would likely be different	28
Staff that are more involved would score higher (more familiar) and vice versa	29
Staff that are more involved would score lower (more critical) and vice versa	30
Respondent ideas about evaluative thinking	31
Study definition of evaluative thinking	32
Evaluative thinking as a social process	33
How does evaluative thinking show up in your/department's work?	34
Reflective practice is/not practiced/examples	35
Assumptions/positionality is/not practiced/examples	36
Systematically collected evidence is/not practiced/examples	37
How does evaluative thinking differ in dept vs. rest of org?	38
Professional development conducted/examples	39
Seeking/using grants to support evaluative efforts/training	40
Respondents ideas about psychological safety	41
Maslow's hierarchy invoked in discussing psychological safety	42

Table 13 (continued)

Code Description	Code
Staff turnover/org stability influencing psychological safety	44
COVID influence work/psychological safety	45
Relationship between psychological safety and evaluative thinking/evaluation culture	46
Evaluative thinking related to process	47
Learning is risky	48
How DEI influences evaluative thinking/evaluation culture (esp. assumptions/positionality)	49
Work with university partnerships	50
How org evaluation is characterized in accreditation documents	51
Evaluation/audience research associated with master/strategic planning	52
Development/marketing/other departments asking for/driving evaluation efforts	53
Felt pressured/inspired by AZA accreditation process/peers to improve evaluation efforts	54
Being open to ideas/input part of personal practice/values	55
Leadership is risk averse	56
Work of internal/program evaluator	57
Work with external evaluators limited to director or small number of staff*	58
Evaluation efforts elsewhere in organization	59
University partnerships often with students, treated as one-offs	60
Work with evaluators has lessened fear of or apprehension about evaluation/built trust	61
Having an internal evaluator/liaison made working with external evaluators easier/less intimidating	62
How do you improve your program holistically	63
Evaluation culture/knowledge/value/use higher in programming/edu department	64
Broadscale training/work with evaluators across department led to improved evaluative thinking practices	65
Informal, team-building-style activities (article clubs, etc.)	66
Team member personalities can contribute to or detract from psychological safety	67
Staff don't see leadership as open to ideas or feedback	68
Contract evaluators work broadly with staff and community	69
How are reports handled/who sees them?	70
Audience research evaluation conducted to meet tax/government requirements	71
Accessibility assessment	72
Provides voice for staff	73
Worked with leaders that have strong evaluation values	74
Leadership support is implied or tacit	75
Encourage trust/safety by helping staff make decision (which involves risk taking)	76
Psychological safety in department vs. org	77
Risk more available to education staff because stakes are lower	78
Internal staff with evaluation experience	79
History/rationale of internal evaluation capacity development at org/in industry	80
Impetus for evaluation/data-drive decision making influenced by science identity of org	81
Staff that don't know the history of evaluation at org might score lower because they don't know how far the org has progressed	82
Professional development through action research	83
Internal evaluators providing formal/informal professional development	84
Develop evaluation capacity by participating in evaluations	85

Table 13 (continued)

Code Description	Code
Internal evaluators facilitate external evaluators working with staff	86
Safety linked to concern that leadership does/not value education work/staff	87
Work of internal audience researchers	88
Challenges in utilizing results of evaluations	89
Co-design/co-creation	90
Balance of desire to do more evaluation and need to finish projects (exhibits)	91
Staff accountability as a contribution to psychological safety	92
ROI as part of the evaluation process	93
Leadership not supportive of evaluation efforts	94
Team is risk-averse (due to previous bad experiences with leadership)	95
Siloing at org diminishes psychological safety (or affects other aspects of evaluation culture)	96
Description of evaluation efforts from accreditation application	97
Use evaluation specifically to improve programs	98
How does evaluation work have impact on broader industry?	99
Leadership modeling risk taking/making mistakes	100
Are scores influenced because respondents don't know what they don't know?	101
Unrelated/irrelevant	999

Following the completion of this study, a short report will be provided to the education director of each case study institution that will include the industry-wide average scores for each dimension of the ROLE instrument and the overall evaluation culture score, the director’s scores, and some notes on observations and trends based on the survey results and case study interviews. These results will have value to the education directors as they work to improve the evaluation cultures within their workgroups and institutions.

Data Analysis

The data from ROLE instrument were analyzed using SPSS (version 25).

Phase one data. Measures of central tendency were calculated for all demographic data to help understand the nature of the sample and how it compared to the broader AZA community. Nonparametric chi-square tests were used to compare the study sample to the AZA population. For completed surveys, mean scores (and other descriptive statistics) were calculated for each of the dimensions (organizational culture, leadership, systems and structures, communication, teams, and evaluation) measured by the ROLE instrument. Two methods were explored to create an overall evaluation culture score (used to identify potential case study institutions). One option used an average of all 50 item responses. A second option used an average of the six dimension means. The second option was chosen because of the variability of the number of items in each scale (ranging from 2-20). Ultimately, the method of calculating the overall score did not result in meaningful differences in subsequent analyses. However, some institutions differed in their final rankings on the overall evaluation culture score by one-three positions. Sample-wide mean scores in each dimension and for the overall evaluation culture score were calculated for comparison purposes. Internal consistency reliability was evaluated for each dimension using Cronbach's alpha. Conditions for the independent variable (work with professional evaluators is summarized in Tables 14 and 15. Contingency tables and chi-square tests were used to explore the relationship between institutional demographics and work with professional evaluators. One-way MANOVAs and ANOVAs (as appropriate) were conducted to look for relationships between evaluation culture scores (including dimension means) and both the institutional demographics and work with professional evaluators variables. A multiple regression was conducted to evaluate the combined relationship between the demographics and evaluator variable and evaluation culture scores.

Phase two data. Phase two consisted of nine case study interviews with select Education directors who completed the phase one survey. Recorded interviews were auto-transcribed and reviewed for each dimension of the survey and the demographic responses to develop a deeper understanding of the evaluation culture at each case study site. Interview content was organized, coded and sorted according to Ose (2016) as described previously.

An emergent construct: Psychological safety. In the course of data analysis, no relationships were revealed between either work with professional evaluators or institutional demographics and evaluation culture. A series of exploratory factor analyses were conducted on the 50 items from the ROLE instrument to find an alternative explanation for the variance in scores. An initial attempt used principal axis factoring with a promax rotation. Using Kaiser's criterion (i.e., eigenvalues $\geq$ 1.0), 13 factors were identified that explained 73% of the variance in the scores. After reviewing the pattern matrix for items that loaded on each factor (>.40), no coherent explanation emerged that explained the variance. As there are six subscales identified in the original instrument, a subsequent exploratory factor analysis was conducted with six forced factors. These did not align with the subscales and did not offer a coherent explanation of the variance. Additional solutions were explored with minimum eigenvalues restricted to 1.5 and 2.0. At 2.0, a solution set of four factors emerged. When reviewing the items that loaded ($\geq$.36) in the pattern matrix along each factor a potential explanation developed. One factor aligned with the evaluation subscale of the original instrument, one factor aligned with the study definition of evaluative thinking, and the final two factors aligned with two aspects of a new construct consistent with the concept of psychological safety (with leadership, and among the team).

Table 14

Categories for Evaluation Resource Data

Internal Evaluators	External Evaluators	Trained (Non-evaluator) Staff
Program evaluators	Contract/consultant evaluators	Yes
Audience/soc. science researchers	Audience/market researchers	No/Not sure
None	University Partnerships	
	Other	
	None	

Note. Internal, non-evaluator staff with training or experience comparable to a professional evaluator (university degree or professional certification, several years of experience).

Table 15

Conditions for Work with Professional Evaluators Variable

Condition	Internal Evaluators	External Evaluators	Trained Staff
1	No	No	No
2	Yes	No	No
3	Yes	Yes	No
4	Yes	Yes	Yes
5	No	Yes	Yes
6	No	No	Yes

Note. Internal evaluators indicates presence of internal program evaluators or researchers. External evaluators indicates work with contractors, external audience researchers, or university partnerships. Trained staff indicates presence of internal, non-evaluator staff with training or experience comparable to a professional evaluator (university degree or professional certification, several years of experience).

With these new factors in mind (evaluation, evaluative thinking, leadership-associated psychological safety, and team-associated psychological safety), an identical set of statistical analysis were conducted replacing the dimension means with the new factors. Means (and other measures of central tendency) and Cronbach's alpha were conducted for the new subscales and compared to institutional demographics and work with professional evaluators through one-way ANOVA/MANOVAs. A new overall evaluation culture score was calculated by obtaining the average of the four factor means and similarly compared.

Synthesis. To answer the research question regarding the relationship between differences in evaluation culture and differences in the extent to which organizations work with evaluators, the results from the analyses of variance were reviewed to identify relationships between interactions with evaluators and evaluation culture scores. The results of the factor analysis and subsequent statistical comparisons involving the new construct of psychological safety were reviewed to determine what insight it might provide in accounting for variance in evaluation culture scores. For selected case studies in each of the three evaluation culture tiers (weak, moderate, and strong), the director interviews were used to: a) clarify the nature of staff and institutional interactions with evaluators, b) understand any relationships that emerged between other institutional characteristics and evaluation culture scores, c) better understand organizational context, and d) explore the emergent construct of psychological safety. By synthesizing data from these sources, a judgement was made regarding the study hypothesis and the possible explanations were put forth for the relationships that did or did not emerge.

Chapter Four: Results

This section reports the results of the study related to the demographics of the responding institutions, trends in organizational work with professional evaluators, the dimension and overall evaluation culture scores from the modified ROLE instrument, emergent themes from the phase two case study interviews, as well as some preliminary findings related to another emergent construct, psychological safety (for further discussion in Chapter Five).

Phase One Data

Phase one consisted of the data related to the modified ROLE survey instrument.

Respondents. At the onset of this study, there were 233 institutions accredited by the Association of Zoos and Aquariums (AZA). The decision was made to delimit participation to U.S.-based institutions to minimize cultural differences (legal, language, professional norms). This restricted the study population to 215 institutions. Contact information was secured for 206 education directors. At the close of the survey, 119 responses had been submitted. After eliminating duplicate and incomplete entries, the final responses totaled 100 institutions, representing 49% of the invited participants (100/206) and 47% of the delimited study population (100/215 U.S.-based AZA-accredited zoos and aquariums). This included 65 zoos, 23 aquariums, and 12 related institutions (safari parks, museums, etc.) from 38 states and the District of Columbia. California, Texas, and Florida were the most represented with 12, 8, and 6 institutions, respectively. There were 15 states represented by a single institution and 12 states did not have a responding institution.

Governance. Governance reflects the operating authority for the institution. Choices in the survey were: for-profit, non-profit, and public (government/municipal). Governance of the sample compared to the population is shown in Figure 2. A chi-square test was conducted to evaluate the representativeness of the sample. The make-up of the sample was found to be different from the population, $\chi^2(2, N = 100) = 9.3, p = .01$, with non-profit organizations represented more significantly in the sample. All comparison statistics for institutional demographics come from AZA's 2018 Benchmarking Reports (Association of Zoos & Aquariums, 2018a).

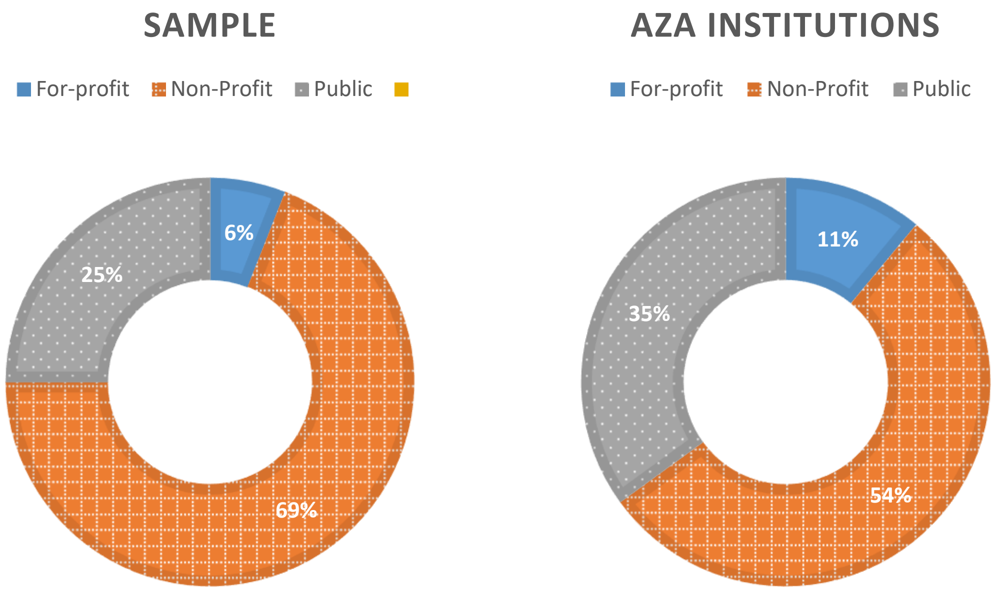


Figure 3. Percentages of governance types among insitutions in the study sample vs AZA member insitutitons.

Operating budget. The size of an institution can be viewed from a number of perspectives. This study has chosen to use institutional budget and annual attendance as estimates of size. Number of employees could be another approach, but these statistics are not as readily available. Information related to budget categories for the sample compared to the population is shown in Figure 3. A chi-square test suggested no significant difference between the sample and population, $\chi^2(3, N = 100) = 4.7$, $p = .195$.

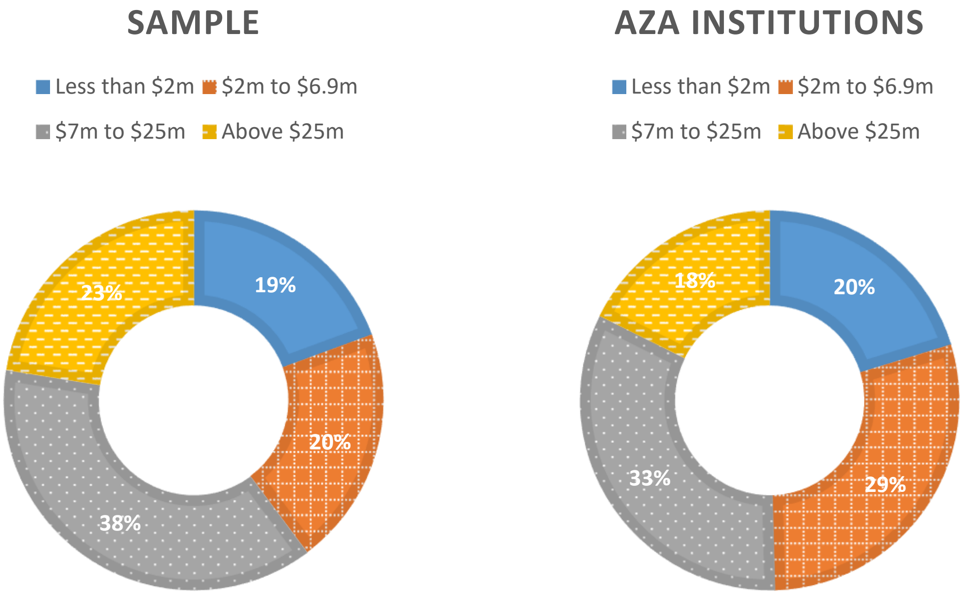


Figure 4. Percentages of annual operating budget categories among insitutions in the study sample vs AZA member insitutitons.

Annual attendance. Annual attendance was included because it represents a different perspective on size. Some free institutions may serve very large audiences, even with smaller revenue and expense budgets. Information related to annual attendance categories for the sample compared to the population is shown in Figure 4. Chi-square testing found no significant difference between the sample and population, $\chi^2(3, N = 100) = 2.9$, $p = .414$.

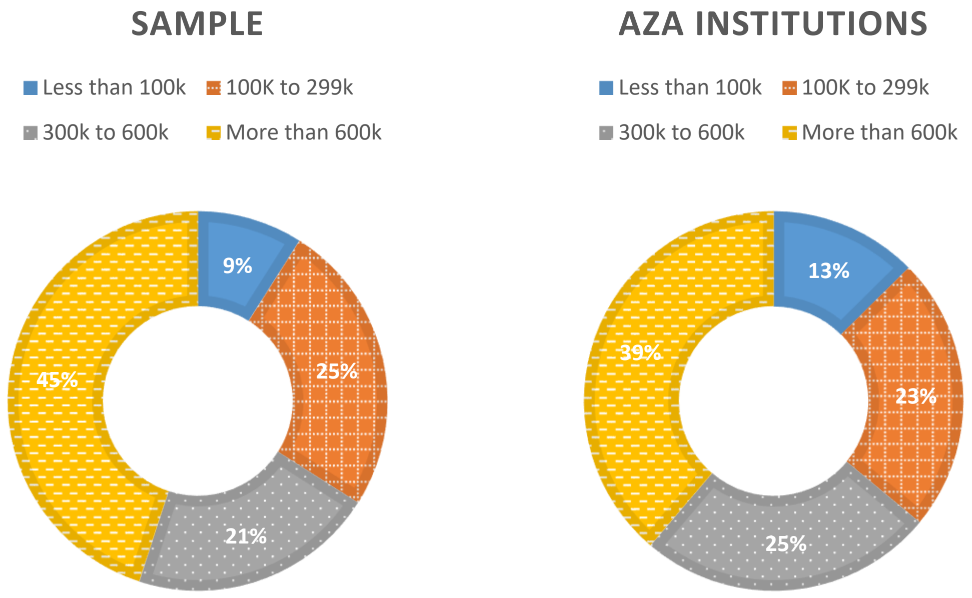


Figure 5. Percentages of annual attendance categories among insitutions in the study sample vs AZA member insitutitons.

Departments and titles. The terms *education* or *learning* were used in 90% of the department titles shared by respondents. Conservation was present in 19%. A total of 40% were designated as directors in their job titles. Curator, manager, and vice-president were also common (19%, 16%, and 16%, respectively).

Work with professional evaluators. In this study, a professional evaluator is an individual with formal training or education in evaluation and/or several years of work experience in an evaluative function. Distinctions were made between professional evaluators who worked for an institution (i.e., internal evaluators) and those who were not employees but who worked with an organization on a temporary or project basis (i.e., external evaluators). A further distinction was made between internal staff whose job was evaluation (internal evaluators) and staff working in other roles that have comparable training or experience (i.e., trained internal staff). See Table 14 in Chapter Three for the categories associated with each.

Internal evaluators. Just 21% of institutions indicated internal evaluation staff of some kind. This included 15% who indicated program evaluators, specifically; 12% who indicated audience researchers, and 2% who noted social science researchers. A majority of these staff worked in the director's home department (62%), but some indicated staff, especially audience researchers, worked in other departments (e.g., marketing, exhibits). There were 20 respondents who indicated the FTE associated with their internal evaluation staff. It ranged from 0.25 to 5 with a mode of 1 ($n = 10$).

External evaluators. Most respondents (90%) indicated some work with external evaluators, including contract evaluators, audience researchers, and university partnerships. Figure 5 shows the frequency with which each institution indicated they work with each external evaluation resource. Contract evaluators, commonly associated with grant and other special projects, and university partnerships were most frequently mentioned (by 70% and 67% of respondents, respectively). External audience researchers were also mentioned by 56% of the respondents. The most common cadence was every few years, which was twice as frequently chosen as the next most common response, several times a year (which could be as little as twice or as many as 10 times). Very few indicated work with external evaluators as frequent as monthly or weekly.

Trained internal staff. Finally, 38% of respondents indicated the presence of internal staff with training or experience comparable to a professional evaluator. Conversely, 49% indicated no such staff, and 13% were unsure. Most indicated that these staff work with their peers in all phases of the evaluation process, as available.

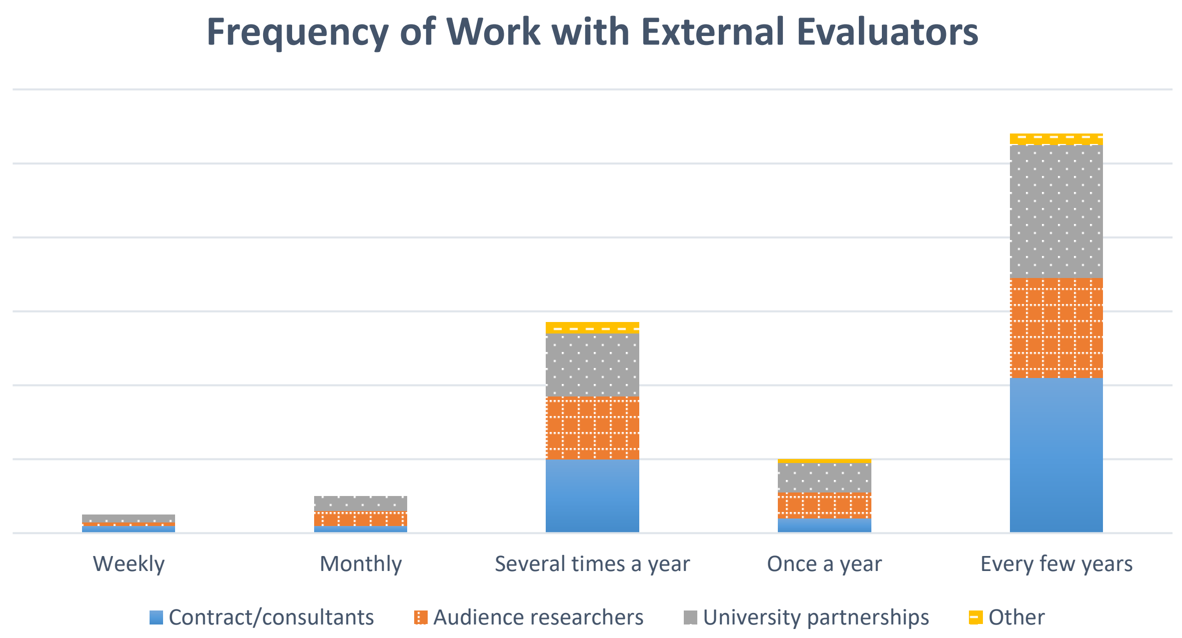


Figure 6. Number of mentions for the frequency of work with each category of external evaluators in the sample. Respondents selected a frequency for each category.

All but one institution with internal evaluation staff also worked with external evaluators. There were 12 of the 20 institutions that indicated work with both internal and external evaluators further indicated the presence of internal non-evaluator staff with comparable training/experience. A small number of institutions (9%) indicated neither internal evaluation staff nor work with external evaluators. Of these, only three indicated the presence of trained internal staff (see Table 16).

Relationship between demographics and work with evaluators. The relationships between institutional demographics and work with professional evaluators were reviewed to understand any confounding influence on the investment in evaluation resources.

Table 16

Conditions for Evaluation Resource Data in the Sample (N = 100)

Internal Evaluators	External Evaluators	Trained (Non-evaluator) Staff	*n*
No	Yes	No	47
No	Yes	Yes	23
Yes	Yes	Yes	12
Yes	Yes	No	8
No	No	No	6
No	No	Yes	3
Yes	No	No	1

Note. Internal, non-evaluator staff with training or experience comparable to a professional evaluator (university degree or professional certification, several years of experience).

Governance. Contingency tables were constructed to examine the relationship between governance and work with professional evaluators. The three governance conditions (for-profit, non-profit, and public) were evaluated against the presence or absence of internal evaluation staff and the positive or negative indication of work with external evaluators (see Table 17). A chi-square test indicated no relationships between conditions for professional evaluators, χ^2(6, *N* = 100) = 8.739, *p* = .19. Governance also had no statistical association with the presence of internal staff with equivalent evaluation training/experience, χ^2(4, *N* = 100) = 2.175, *p* = .70.

Table 17

Work with Professional Evaluators by Institutional Governance

Condition	For-profit	Non-profit	Public	Total (row)
	%	%	%	*n*
No work with professional evaluators	33.3	5.8	12.0	9
Internal indicated, external not indicated	0.0	0.0	4.0	1
Internal not indicated, external indicated	50.0	72.5	68.0	70
Both internal and external indicated	16.7	21.7	16.0	20
Total (column) *n*	6	69	25	100

Institutional size. Similar tests were conducted for the four categories (small, medium, medium-large, large) of each of the two institutional demographic variables associated with size: operating budget and annual attendance (see Table 17 and Table 18). Both size-related variables showed a significant relation with organizational work with professional evaluators. For operational budget: χ2(9, n = 98) = 37.35, *p* = .00, and for annual attendance, χ2(9, N = 100) = 17.79, *p* = .04. Neither showed a significant relation with the presence of internal staff with equivalent evaluation training/experience, χ2(6, *N* = 98) = 3.906, *p* = .69 and χ2(6, *N* = 100) = 5.552, *p* = .48. The results were similar if not sure was grouped with no when considering the question of internal staff with evaluation training/experience. That larger organizations were associated with more investment in evaluation resources was an expected finding.

Table 18

Work with Professional Evaluators Versus Operating Budget

Condition	Small %	Medium %	Med-large %	Large %	Total (row) *n*
No work with professional evaluators	21.1	5.0	10.8	0.0	9
Internal indicated, external not indicated	0.0	5.0	0.0	0.0	1
Internal not indicated, external indicated	78.9	85.0	75.7	40.9	69
Both internal and external indicated	0.0	5.0	13.5	59.1	19
Totals (column) *n*	19	20	37	22	98

Evaluation culture and dimension scores. The original ROLE instrument consisted of 74 items in six dimensions (subscales), organizational culture, leadership, systems and structures, communication, teams, and evaluation. The instrument was shortened after feedback from reviewers to 50 items with items from each dimension. Each item was scored on a scale from 0-100 (strongly disagree to strongly agree).

Table 19

Work with Professional Evaluators by Annual Attendance

Condition	Small	Medium	Med-large	Large	Total (row)
	%	%	%	%	*n*
No work with professional evaluators	11.1	16.0	9.5	4.4	9
Internal indicated, external not indicated	0.0	4.0	0.0	0.0	1
Internal not indicated, external indicated	77.8	80.0	76.2	60.0	70
Both internal and external indicated	11.1	0.0	14.3	35.6	20
Totals (column) *n*	9	25	21	45	100

Organizational culture. The organizational culture dimension contained 20 items addressing topics of collaboration, problem-solving, risk-taking, and decision-making. The mean score on the organizational culture dimension was 79.96 with a standard deviation of 9.5, indicating relatively strong agreement by respondents that their organization culture was healthy by the measure of these items. Scores were normally distributed. Reliability for this dimension was calculated for the 20 items using Cronbach's alpha and determined to be strong (α = .89). Descriptive statistics for each dimension and the overall evaluation culture score can be found in Table 20.

Leadership. The leadership dimension contained eight items addressing management views on and support for evaluation. The mean score for the leadership dimension was 80.64 with a standard deviation of 11.2 points with no significant departures from normality. Reliability was strong (α = .82).

Table 20

Descriptive Statistics for Overall Evaluation Scores and Dimensions

Scale and subscales	Min	Max	Mean	*SD*	Skewness	Kurtosis
Organizational Culture	48.60	96.50	79.96	9.51	-0.54	0.80
Leadership	46.25	100	80.64	11.20	-0.77	0.80
Systems	33.57	97.14	74.92	13.66	-0.62	0.25
Communication	10.50	100	60.79	20.39	-0.55	-0.09
Teams	39.00	100	74.74	13.75	-0.19	-0.67
Evaluation	46.25	98.75	77.68	11.98	-0.61	0.14
Overall Evaluation Culture	46.56	96.73	74.79	10.50	-0.26	-0.23

Note. *N* = 100. Scores on a 0-100 scale.

Systems and structures. The systems and structures dimension consisted of seven items with questions about organizational processes and policies that encourage learning and/or evaluation efforts. The mean for this dimension was 74.92 with a standard deviation of 13.7. Scores were normally distributed. Reliability was acceptable (α = .78).

Communication. The communication dimension was reduced from the original instrument to two items concerning information/data collection and past records of evaluation/learning. The small number of items likely contributed to the high standard deviation (*SD* = 20.4) and moderate reliability (α = .61). The mean for the communication dimension was also quite a bit lower than other dimension means 60.79, drawn down by a very low mean score on the second item (*M* = 49.36) related to institutional records of past change efforts.

Teams. There were five items in the teams dimension that address the functioning and norms of teams within the department. Scores were normally distributed with the mean score of 74.74 and a standard deviation of 13.7. Reliability was moderate (α = .65). Communications with some respondents from small institutions (including case study interviewees) indicated difficulty responding to these items as their department might have only one to two staff. Some chose to answer with their organization in mind rather than their department, which diverges from the framing instructions, but is not likely to have significantly altered the intent of the questions.

Evaluation. Eight items made up the final dimension, covering the use, interest, and support for evaluation in the department. The mean for the evaluation dimension was 77.68 with a standard deviation of 12.0 and solid reliability (α = .71) with no departures from normality.

Evaluation culture. An overall evaluation culture score was calculated by averaging the means for each dimension. The resulting score is on the same scale as the individual items (0-100). The mean evaluation culture score among the 100 institutional respondents was normally distributed with a mean of 74.79 and a standard deviation of 10.5. Reliability was strong (α = .85). A similar score was calculated by taking the mean of all 50 items from the instrument. The mean for this score was higher (M = 77.71) with a slightly smaller standard deviation (SD = 9.6) and a higher Cronbach's alpha (α = .94), but there was concern that the score would be unduly influenced by the larger organizational culture dimension, which had more than twice the number of items of any other subscale. Pearson correlations were calculated for the six dimensions that make up the overall evaluation culture score. Correlations ranged from .39 to .76 and were each significant at the $p < .01$ level See Appendix J, Table J1, for the correlation matrix.

Relationships between demographics and evaluation culture scores. Just as is it expected to see larger institutions with more resources to invest in evaluation, it might also be expected that size or budget might be related to an organization's evaluation culture, regardless of their work with evaluators. Institutional demographics were therefore explored for potential relationships to evaluation culture scores.

Governance. A one-way ANOVA showed no relationship between governance and overall evaluation scores, *F*(2, 97) = 0.02, *p* = .98. A one-way MANOVA conducted to examine the six dimension means showed no significant relationships, *F*(12, 184) = 1.50, *p* = .127; Wilk's Λ = 0.83, partial η^2 = .09.

Institutional size. One-way ANOVAs were conducted similarly for the institutional size variables of operating budget and annual attendance. Neither demonstrated a significant relationship with the overall evaluation culture score, *F*(3, 94) = 0.356, *p* = .79 and *F*(3, 96) = 2.123, *p* = .10. No relationships were found between the dimensions and budget, *F*(18, 252) = 1.32, *p* = .174; Wilk's Λ = 0.78, partial η^2 = .08. However, there were two significant relationships revealed in MANOVA tests of annual attendance by dimensions means. A one-way ANOVA for organizational culture scores showed a significant relationship with annual attendance, *F*(3, 96) = 3.07, *p* = .03. Bonferroni post hoc tests showed a difference between the medium-large (*M* = 84.50, *SD* = 8.12) and large categories (*M* = 77.76, *SD* = 8.65) with a relatively large effect size, *d* = .79. A similar relationship was found in a one-way ANOVA between attendance and the systems and structures dimension, *F*(3, 96) = 3.37, *p* = .02 with a significant difference seen again between the medium-large (*M* = 81.99, *SD* = 10.25) and large categories (*M* = 71.03, *SD* = 13.56) and a similarly large effect size, *d* = .87. No other significant relationships were identified.

Relationships between work with evaluators and evaluation culture. The hypothesis of this study stated that institutions invested in internal evaluators will have stronger evaluation cultures. It was also hypothesized institutions working with both internal and external evaluators would have even stronger evaluation cultures (though the results may not be significant). The second hypothesis was not testable in this sample since all institutions with internal evaluation resources (save one) also indicated work with external evaluators.

Internal evaluators. A one-way ANOVA looking at the relationship between overall evaluation culture score and the presence or absence of internal evaluation staff found no significant difference, $F(1, 98) = 1.16$, $p = .29$. An ANOVA was also conducted to investigate individual internal evaluator categories (program evaluations, researcher, evaluators and researchers, and no internal evaluators). No significant difference between group means was detected, $F(3, 96) = 1.16$, $p = .33$. A one-way MANOVA was conducted to examine whether the presence/absence of internal evaluation staff was related to any of the six dimensional means. No significant relationships were detected, $F(6, 93) = 1.92$, $p = .09$; Wilk's $\Lambda = 0.89$, partial $\eta^2 = .11$.

External evaluators. A one-way ANOVA for presence/absence of external evaluators similarly returned no significant differences in evaluation culture, $F(1, 98) = 0.71$, $p = .40$. A one-way MANOVA for the dimensional means also showed no significant differences, $F(6, 93) = 1.75$, $p = .117$; Wilk's $\Lambda = 0.90$, partial $\eta^2 = .10$. Frequency of work with external evaluators could not be evaluated because of too few cases in multiple conditions.

Trained internal staff. Like the above cases, a one-way ANOVA evaluating the relationship between the presence/absence of trained (non-evaluator) internal staff and overall evaluation culture score returned no significant results, $F(2, 97) = 1.50$, $p = .23$. A MANOVA for the dimensional means again showed no significant differences, $F(6, 93) = 1.11$, $p = .361$; Wilk's $\Lambda = 0.93$, partial $\eta^2 = .07$.

Other comparisons. A one-way ANOVA was conducted comparing the overall evaluation culture scores against the seven presence/absence conditions for professional evaluators and trained internal staff (see Table 16) showed no significant differences between means, $F(6, 93) = 1.32$, $p = .26$. Results for a one-way MANOVA comparing the conditions to the six dimensional means were similar, $F(36, 389) = 1.17$, $p = .24$.

A multiple regression was designed to predict overall evaluation culture score according to institutional governance, annual budget, and the seven presence/absence cases for work with professional evaluators. Annual attendance was excluded, because it was highly correlated with institutional budget. The result accounted for just 10% of the variance, $R^2 = .10$. See Table 21 for results.

Exploratory factor analysis. Since no meaningful relationships were identified when looking at institutional demographics or work with professional evaluators, a principal axis exploratory factor analysis with promax rotation was conducted to look for an alternative explanation of the variance in the 50 item ROLE scores. Four new dimensions were identified from the 50 items which aligned with two existing study constructs and two emergent constructs. There were 42 items that loaded on one of the four factors ($\geq .36$ in the pattern matrix, see Tables J2 and J3 in Appendix J for the pattern and structure matrices). Factors correlations ranged from .45-54 (see Appendix J, Table J4). (A full discussion of the rationale associated with the identification of the four constructs can be found in Chapter Five.)

New dimension: Evaluative thinking. There were 19 items associated with factor one. Items from all six original subscales were included and collectively provide a strong analog for the study construct of evaluative thinking (a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making). The mean for the new evaluative thinking dimension was 71.16 with a standard deviation of 12.75 and strong reliability (α = .91). Scores were normally distributed. See Table 22 for a summary of statistics, including skewness and kurtosis.

Table 21

Results of Multiple Regression on Overall Evaluation Culture

Variable	*B*	*SE B*	β	*p*
For-profit[a]	-7.68	5.78	-.15	.19
Public[a]	-0.16	2.59	-.01	.95
Budget	0.57	1.23	.06	.64
Evaluator condition[b]				
NoIn--NoEx--NoTrained[c]	7.90	4.82	.18	.11
YesIn--NoEx--NoTrained	15.54	10.58	.15	.15
YesIn--YesEx--NoTrained	3.81	4.31	.10	.38
YesIn--YesEx--YesTrained	1.53	3.75	.05	.69
NoIn--YesEx--YesTrained	5.89	2.68	.24	.03
NoIn--NoEx--YesTrained	2.49	6.25	.04	.69
$R^2 = .10$				

Note.· [a]Reference group for governance variable is non-profit. [b]Evaluator conditions indicate presence/absence of internal evaluations--external evaluators--trained (non-evaluator) staff. [c]Reference group for *Evaluator condition* is NoIn—YesEx--NoTrained.

New dimension: Evaluation/growth. Seven items were associated with the factor four, including four of the seven items from the original evaluation subscale from the ROLE instrument. The mean for the new evaluation dimension was 82.96 with a standard deviation of 12.04, good reliability (α = .77) and no departures from normality.

New dimension: Team-related psychological safety. Six items loaded on factor three that together addressed elements of team-related psychological safety. The mean for this dimension was 85.00 with a standard deviation of 10.02, a normal distribution, and strong reliability (α =

.85). The association of factors two and three with psychological safety is based on (Edmondson, 1999) and related literature (see Chapter Five).

Table 22

Descriptive Statistics for New Overall Evaluation Scores[1] and Dimensions

New Dimensions (Subscales)	# Items	Min	Max	Mean	*SD*	Skew.	Kurt.	*α*
Evaluative Thinking	19	36.05	95.79	71.16	12.75	-0.39	-0.03	.91
Evaluation	7	39.29	100.00	82.96	12.04	-1.06	1.20	.77
Team-related psych. safety	6	55.50	100.00	85.00	10.02	-0.82	0.28	.85
Leadership-related psych. safety	10	43.90	100.00	82.72	11.00	-0.82	0.58	.86
Total psychological safety	16[a]	48.25	98.13	83.57	9.66	-0.87	0.92	.90
Overall Evaluation Culture	42	46.45	95.71	77.86	9.73	-0.46	0.23	.80

Note. *N* = 100. Scores on a 0-100 scale.
[a]Total psychological safety is a combination of the team-related and leadership-related scales.

New dimension: Leadership-related psychological safety. A total of 10 items loaded on factor two which was associated with leadership-related psychological safety. The mean was 82.72 with a standard deviation of 11.00, a normal distribution, and strong reliability (α = .86).

Total psychological safety. The 16 items that loaded on factors two and three together were used to create a total psychological safety score. The mean for total psychological safety was 83.57 with a standard deviation of 9.66, a normal distribution, and strong reliability (α = .90). The leadership- and team-related dimensions were positively correlated [$r(100) = .62$, $p = .00$)].

A new way to calculate evaluation culture. To explore whether the subscales associated with the new factors represented a new way to think about an overall evaluation culture score, the four means were averaged and used to evaluate potential relationships between institutional demographics and work with professional evaluators. The new overall mean for overall evaluation culture score was 77.86 with a standard deviation of 9.73 and strong reliability (α = .80). Scores were normally distributed.

Relationships between demographics, work with evaluators, and new evaluation culture scores. Relationships between demographics, work with professional evaluators and the new overall evaluation culture scores were explored, although the conclusions are limited by the fact that the instrument was not designed to measure psychological safety specifically. Results were similar to those conducted around the original overall evaluation culture score.

Institutional demographics. No significant relationships were determined when one-way ANOVAs were conducted for the new overall evaluation culture score versus governance, operational budget, and annual attendance. A one-way MANOVA comparing each institutional demographic to the new factors (evaluative thinking, evaluation, team-related psychological safety, leadership-related psychological safety, and total psychological safety) similarly showed no significant relationships. See Tables J5 and J6 in Appendix J for full results from ANOVA and MANOVA testing.

Work with professional evaluators. Similar tests were conducted with the new overall evaluation culture score swapped out for the overall score associated with the original six subscales. Tests for differences between new evaluation culture scores and presence/absence of internal evaluators or the internal evaluator categories (program evaluations, researcher, evaluators and researchers, and no internal evaluators) show no significant relationships. Likewise, separate ANOVA tests comparing presence/absence of external evaluators and comparing presence/absence of trained internal (non-evaluator) staff featured no significant differences.

A one-way ANOVA conducted comparing the overall evaluation culture scores against the seven presence/absence conditions for professional evaluators and trained internal staff showed no significant differences between means, $F(6, 93) = 1.13$, $p = .35$. Results for a one-way MANOVA comparing the new factor means, however, were significant, $F(25, 332) = 3.23$, $p < .01$; Wilk's $\Lambda = 0.45$, partial $\eta^2 = .15$ (with one condition excluded because it returned only a single case). Between subjects effects showed a significant result for the comparison of professional evaluators on team-related psychological safety, $F(5, 93) = 2.34$, $p = .05$, but post hoc tests did not identify specific groups where this difference was significant. When looking at one-way MANOVAs for the presence/absence of internal evaluators and external evaluators individually, significant results were found in each case—$F(5, 94) = 2.56$, $p = .03$ and $F(5, 94) = 4.13$, $p < .01$—but not when looking at presence/absence of trained internal staff individually, $F(5, 94) = 1.05$, $p = .39$. In each case, means for team-related psychological safety were higher in the absence of evaluators. A multiple regression was conducted to predict the new overall culture score according to institutional governance, annual budget, and the seven presence/absence cases for work with professional evaluators (annual attendance excluded again because it was highly correlated with institutional budget). The result was very similar, accounting for just 10% of the variance, $R^2 = .10$. See Table J7 in Appendix J for full results.

Phase Two Data

Phase two consisted of nine follow-up, case study interviews. Respondents were ranked by their overall evaluation culture scores (using the original measure), which ranged from 96.73

to 46.56. Three tiers were created representing *strong*, *moderate*, and *weak* evaluation culture scores (relative to the sample) by dividing the sample into thirds. The strong tier included scores ranging from 96.73 to 80.18, the moderate tier from 80.11 to 71.87, and the weak tier from 71.65 to 46.56. Three case studies were selected in each of the three tiers. Interviews were transcribed, organized, coded, and sorted by themes. The final sorted codes are found in Table 23. Top-level themes included: Organizational circumstance, work with professional evaluators, evaluation culture, evaluative thinking, and psychological safety. These top-level themes were driven primarily by the interview protocol, which was, in turn, developed to reflect the core constructs of interest in the study. All case-study interview quotes are referenced with the anonymized institution label (A-I) and an indication of its tier (Strong, Moderate, or Weak).

Organizational circumstance. This initial theme centered around the individual director (including their background in evaluation), their department and its structure, and how the evaluation efforts have been characterized in accreditation documents. A typical response was:

> So, there's the education curator, which is me. And then we have three different sections. We have outreach, public programming, which is all the programs that are included with your admission ticket, and then special activities which are programs that require an additional fee, as well as our volunteer coordinator who oversees about 86 active volunteers. Our Special Activities department are the ones who do the birthday parties overnights camps, so on and so forth. Public Programs are the ones who do tank talks, live animal encounters and so on. And

> then [staff person] does outreach, which normally would include going off-site, as well as our distance learning. (Institution G: Weak)

Interviews reviewed or requested the accreditation documents related to their evaluation activities (a requirement of the AZA accreditation application, section EI-9). Some had provided these in advance, others were discussed in the interview and provided later. As the accreditation cycle is five years, some had not been the authors of the previous application, were less familiar with it, or felt it no longer reflected their activities accurately.

> Um, I mean to be honest, I'm not entirely sure exactly how we would have worded it. It was definitely on the last accreditation, which was three years ago, I think. We were challenged to do some better, more thorough evaluation of our programming. We had been using a tool that had been created before I started there and it was, you know, it was working okay, so we just kept doing it. Since then we've tried a few different versions of the tool and now my coworker, our coordinator, she's been attending some, just a three or four session course, developed or presented by the [local environmental educators association] and so she's a little bit more about evaluation and it's been helping her to create some better evaluations that we're going to employ, particularly for our virtual programs that we're putting into place, as well as when we go back to in-person. They're just more thorough and more focused on what our, what we want to know for outcomes are, you know, hitting whereas before it was like, did you like it. (Institution B: Strong)

> Honestly, I didn't feel that one out, my predecessor did so I mean, we’re coming up on another accreditation cycle. We were supposed to actually be accredited this year and then due to COVID it got moved to next year. So, I have just pulled all the AZA standards out and I'm starting to work on that now. (Institution G: Weak)

As suggested in the quote above by Institution B, some found accreditation as a motivator to improve their practice.

> I think when we had our AZA [accreditation], that was when it kind of became clear to [our director] that we needed to beef up that part of what we were doing. I think we're by far, you know, trying to better our evaluations, make sure they're always a part of programs and the feedback involved with it, much more so than anybody else [in their institution]. (Institution B: Strong)

> . . . so I think it would feel a gap, and I think one of the reasons for that is the AZA association . . . I mean, you see the work that other folks are doing or the fact that there are people with these positions available in institutions and so . . . [y]ou would be wondering, why aren't we doing those things or why don't we even talk about that? (Institution D: Moderate)

Directors spoke about their training and experience in evaluation, which was largely informal. Some had courses during graduate work or did evaluation-related projects within other courses, but most cited their work doing evaluation activities as an educator or participating in evaluation activities led by professional evaluators. At least one also mentioned the positive influence of previous leaders.

It's mostly informal. I have done things like, our local [university] has an informal science education degree that they offer, and as part of that they did a semester long training for people in the field, so I participated in a program with that. My Masters is in museum education and so evaluation was a component of that. And then just doing work related to all the projects I've been a part of. So before I was at the [institution], I was at [previous institution], and we were in the process of opening a bunch of new exhibits for the first time in 30 some years. . . and so evaluation was a big part of that process, obviously, as we went through the entire thing, but it's mostly informal. I mean, I haven't had really in-depth training related to it.
(Institution A: Weak)

I haven't gone through any training specifically, it has been more going to some of the sessions at AZA, also seen round tables. And then I was very fortunate, when I first started in this field, my boss, the curator, he was getting his master's and he was doing his thesis on evaluation. (Institution I: Strong)

. . . working at [previous institution], working with [previous supervisor] and [previous supervisor] with their backgrounds and their focus and then [previous institution] obviously, [previous supervisor] and [evaluator] and now I see [new evaluator] is on their team there. So they've got, I think maybe three people now on there. So they're building a pretty robust internal evaluation culture. (Institution D: Moderate)

Work with professional evaluators. The interviews provided an opportunity to understand the context and details of their responses around their department's and institution's work with professional evaluators and trained internal (non-evaluator) staff. Sub-themes included the nature of their internal evaluation staff and their work (for Institutions C, E and F) the details of their work with external evaluators, evaluation efforts in other parts of their organization, and the potential influence of evaluators on leadership or staff views of evaluation. Directors shared that, in addition to their own professional motivations and requirements of accreditation, sometimes strategic planning was the driver of evaluation efforts.

> We're putting together a conservation strategic plan at the moment, and just in the last stages of that, that's something new that hasn't existed. And so I think everyone is actually going through this process in their own way, and we're all pulled into it in bits and pieces as we work together on those things . . . a lot of different divisions are moving in this direction and doing a lot of work to be more evaluative. (Institution A: Weak)
>
> So, we have a new leader that started a year ago . . . and so, we're in the process of writing a new strategic plan. And with that, [becoming] a much more cause-driven organization. We're also writing these initiatives on how we're going to implement the strategy. So, and in each one of these we're actually thinking very hard about including social sciences, and it's not just market research. Like, we need to know when we're interacting with these teachers on these issues, or community groups, . . . how are we moving the needle? Are we really changing the world around us

> through this work, and if not, how do we make it more impactful? (Institution F: Moderate)

For institutions with internal evaluation staff, directors shared details on how those staff work and the history of the organization that led to their hiring. One institution worked over 30 years to create an entire department of evaluation and research. Another just recently converted an education staffer with a good mind for data to help organize their evaluation efforts. Other, institutions, like the author's home institution, have hired a single professional evaluator to both conduct evaluations and build evaluation capacity. The presence or absence of internal evaluation staff at an institution can be quite variable. Institutions with internal evaluation staff mentioned benefits that included, reductions in anxiety around evaluation efforts, facilitation of work with external evaluators, and opportunities for training and learning for other staff. Institutions also take advantage of staff with experience in evaluation,

> And so, it's not what their job is, but their experience informs what they do so, and that has mostly to do with what their degrees are in and graduate school and stuff like that. So, um, let's see, [colleague] sits in the student education department. My boss . . . has a PhD in various educational and political things, so, she has a lot of experience in evaluation. In my department . . . that has experienced from [their] museum education graduate work, [colleague and colleague]. I know there's a few more scattered around on the different departments. (Institution F: Moderate)

Marketing and development or philanthropy departments are sometimes partners or drivers for evaluation work in that they are motivated to provide data to their stakeholders to justify decision-making.

> [W]e have a solid development person who's focused on grant funding and she's made it abundantly clear that we need to have data to give these people so that they can make a decision. So we can, you know, backup what we're doing. (Institution B: Strong)
>
> [My colleague], who's the director of marketing. We were kind of like co-conspirators in building an evaluation culture by being dangerous. Learning how to do it. It's not just education, but also looking at exhibits, looking at the market research. You know, trying to look across because there's so much convergence . . . the institution is not just the silo of education. We managed to convince the institution that we needed to hire designated staff . . . It was a great opportunity to build professional development for the Education Team, you know, and the value of research, the value of evaluation, which sometimes gets conflated a little bit. [My colleague] and I proposed that we co-manage an evaluator/market/audience researcher and . . . it worked. (Institution E: Strong)

Work with external evaluators was primarily discussed as finite, project-based activities associated with a grant or exhibit. In some cases, the exposure to these external evaluators was limited to one or a few staff. In at least one case, however, an effort was made to expand the work to be more inclusive of staff across the department. Institutions in the case studies rarely worked consistently with the same external evaluators, some limited by contracting guidelines and others by opportunity. Other sub-themes within this category included work with university partnerships and external audience researchers. The limitations of exposure to all evaluators, internal and external, are discussed in Chapter Five as a challenge that may have contributed to the lack of effect seem in the study survey.

Evaluation culture. Interviews discussed their own views of what makes an evaluation culture alongside the study definition, including collecting data and using evaluation specifically to improve programs. Can your evaluation efforts improve your program holistically (not just a program at a time) or even your industry?

> For me, when you say evaluation culture, I immediately jumped to: do people perceive evaluation as either punitive, or looking for problems, or do people understand that evaluation is about constant improvement? And so, when I think about evaluation culture, what I almost I tend to think about is, where's your organization on that continuum? Not necessarily that one side is, “I'm afraid of evaluation” and the other side is, “I fully embrace evaluation,” but the understanding of what the goals are in building evaluation into what you're doing as just a general best practice. (Institution C: Weak)
>
> To me, [the idea of an evaluation culture] says that it's not just about the evaluation, but it's about knowing what people need and being able to change with that need, being able to pivot as needed, and not being stuck in the, in the same way, just because this is the way it's always done. The evaluations are a tool to get us to where we're helping the most people . . . And I think that's where the culture comes in, that always evaluating what we're offering and whether it's worth it, to keep going that way . . . [or] if it doesn't work, we'd go okay we tried, but also, not to continue doing something just because we did try it. And now we've got to

> improve it. You know, it's that nice feeling of being able to fail and not even fail, to have to change. It's basically a scientific thought process that you have to try and try until you actually get to the final outcome that you want, and I love that we have many different methods to do it. (Institution I: Strong)

Interviewees reflected on their responses to survey questions and how they rolled up into dimension and overall evaluation culture means. For the most part, interviewees felt their scores fairly reflected their feelings about their department compared to other staff at their organization and against the broader sample means. There was agreement that staff and leadership valued evaluation conceptually, but that institutional and professional barriers prevented widespread adoption of evaluation practices and the utilization of evaluation results (though pockets of evaluation capacity could exist in departments or organizations).

> I would say [our evaluation culture] is in the forming stage. It is something that, at leadership levels, both within the division and within the organization, there's a recognition of its importance . . . The desire is there and it's . . . it's one of those things where nobody argues whether we should be doing it, or whether it's important, but it's just a matter of how do you? How do you make time for it? (Institution A: Weak)
>
> . . . not for any malice towards evaluation or anything, but going back to [that] money and time piece. It's easy for me to justify, you know, let's get another body to run summer camp, because then it allows us have more kids to come to summer camp. [But if] I need money for an evaluator, well, where does that come from? (Institution D: Moderate)

Directors generally agreed that the evaluation culture of their departments was stronger than other departments in their organization. Directors felt their leadership were generally, but not universally supportive of evaluation, though that support may have been tacit or poorly demonstrated by resources. Three of the nine directors felt their staff would score the survey similarly to them. The other six offered several explanations for why employees might score differently, including involvement and familiarity with current and historical evaluation activities.

> If you haven't personally . . . if you're a part-time employee and you've only been here a year, then you're just less likely to have been involved or remember . . . you're not the person developing the program whereas [a veteran staff person] who'd been there for five years has developed multiple programs . . . wrote several of the assessments . . . I think it would just depend on where you are in the hierarchy of the department and what your role is and how long you've been there to remember all the different pieces. (Institution F: Moderate)

> I think some folks would score it much lower [and] I think others would be like, "Yeah, we're doing great." We are sending out surveys and we're talking to people and we're making changes . . . so I think I think would end up being pretty similar to that on average, but any anyone might be super high or super low compared. (Institution A: Weak)

> I think some of them would [score items differently] because their context is different. I pride myself in being part of creating an evaluation culture at/in our department and our organization, leading that charge. So, I certainly may be more positive about this because I know where we were, right? Other people who came

> on later, it could be like yeah, we can be so much better if blah blah blah happened. (Institution E: Strong)

Evaluative thinking. Similarly, directors discussed their ideas about evaluative thinking compared to the study definition, with questions about evaluative thinking as a social process, “like organizational constructivism.” (Institution A: Weak). Directors could most easily find examples of reflective practice and data collection in their department’s practices, with some attributing reflection as a natural extension of an educator mind.

> We're educators. So, there's a lot of talking that just happens, because most of the public programs team can't help themselves. So, I don't know that there's anything formalized like, you work with you, who works with you and you guys figure this out and talk to each other and then bring it back and share it with everybody. It's more of a hey, I'm working on this project I want someone to help me think through it. Sometimes we team people up on purpose because they need to learn from each other or somebody can guide somebody new to a process or what have you, but there's very little that anyone does by themselves. (Institution F: Moderate)

> I will say the reflective part. Man, I wish we had more time for that. And I just think that I remember [my colleague] did some research when she was [at peer institution] about that part, you know, as educators are doing evaluation and then how much time do we have to reflect on it, and then act on it? [It is] the reflective part . . . I don't know if it's if it's our nature or what we're asked to do. I think it's probably maybe a little more our nature. It’s that we just what we do It's, it's really

> easy to get stuck in the doing the same thing because it gets really good results. (Institution E: Strong)

> I think different elements show up more than others . . . So I think that the idea of this reflective practice is something that is in our divisional culture and trying to have it be something that all of us are doing together. (Institution A: Weak)

Examples of identifying assumptions or examining positionality were influenced by DEI (diversity, equity, and inclusion) work at their institutions.

> There's a ton of the new strategy all about social justice, environmental justice, the interrelationship between that and biodiversity . . . which wouldn't have happened, I think, unless this year happened the way it did. So, it's had good ramifications and it also has helped, I think, further things that had already started. (Institution F: Moderate)

Directors share examples of professional development around some of the skills associated with the study definition, but none were specifically targeted for their ability to develop evaluative thinking, per se. Participating in evaluations was one of the activities mentioned as contributory to the development of better evaluative thinking.

Psychological safety. Though the term was new to most, the ideas associated with psychological safety resonated with all nine case-study interview participants, aligning with another construct with which they were more familiar.

> I mean, I would say so my first educator answer is like Maslow . . . (Institution A: Weak)

> It's Maslow's hierarchy . . . (Institution H: Moderate)

Directors discusses their ideas about what it meant to feel safe at work compared to the Edmondson (1999) definition and its core components, especially taking risks. Directors identified a series of factors that contributed to and detracted from psychological safety in their teams. Positive influences included being open to ideas, facilitating decision-making and staff empowerment, leadership modeling positive behaviors (including risk-taking), and reducing anxiety (especially related to evaluation).

> One of the things from the beginning, very intentionally, has been about trying to help people to make decisions on our team, like our managers, but also help people all down the line to be able to make decisions. (Institution D: Moderate)
>
> Yeah, and I think it's a way to give a voice maybe to teams that, you know, often feel like they don't have a voice, or that it's not always reflected in what is seen in the zoo. (Institution D: Moderate)

Team-related negative contributors included staff turnover and organizational instability (exacerbated by pandemic impacts), toxic staff personalities, and risk aversion related to past negative consequences. Directors mentioned leadership-related negative contributors such as undervaluing education efforts, a lack of openness to staff ideas, and leaders demonstrating risk-averse behavior. Broader contributors were also mentioned, like siloing at the organizational and department level, staff accountability, and COVID-related uncertainty.

> [A]ll of my team is fully aware that I don't care if they try something and it fails. That doesn't bother me a bit because we will learn from it. I will say that probably my team is not a risk-taking team, and I think that honestly has to do with some leadership before me [and a lack of] openness to differences. I will definitely say

> those two would have made that a challenge, leadership-wise. [Low scoring on psychological safety] doesn't surprise me at all. (Institution G: Weak)
>
> I'm thinking about a couple of the leaders that I have on my team and I just happen to know that at least one or two of them are viewed as very black and white, and less open to ideas . . . I think that I've got a diversity on my team of leaders that contribute to psychological safety and those who probably detract from it. So, where, where the challenge lies is finding a way for that to sort of all work together, because I think what probably influenced [lower scores on items related to psychological safety] was, maybe, the openness of leadership for input from staff is one thing that I've heard through the grapevine is not perceived to be present. Of course, I always look at myself and say, geez, I hope that's not me, and maybe some of it is, but I can also look at some of the leaders that I have in place and I could tell you exactly which ones I think, if I had, if I almost had the continuum of psychological safety, which ones would be on one side of perhaps contributing to most of these versus the others that would probably be perceived as doing less so. (Institution C: Weak)

Unsorted. Some coding and statements did not contribute to the development of ideas around study constructs. They are included in an *unsorted/unrelated* category in Table 23, including a *999* code for small talk, asides, and undecipherable snippets from the conversations. These are example excerpts from the nine interviews. Full transcripts have not been included because they contain identifiable information even with institutional and participant identifiers anonymized.

Table 23

Interview Codes Sorted by Theme

Categories and Codes	Codes
Organizational Circumstance	
Position/department description/organizational circumstance	2
How org evaluation is characterized in accreditation documents	51/97
Felt pressured/inspired by AZA accreditation process/peers to improve evaluation efforts	54
Director background/training in evaluation	9
Director training primarily informal/learn on the job	10
Worked with leaders that have strong evaluation values	74
Work with Professional Evaluators	
Professional evaluator definition	3
Ways respondent/department/org has worked with professional evaluators	4
Evaluation/audience research associated with master/strategic planning	52
How are reports handled/who sees them?	70
Internal Evaluators/staff	
Work of internal/program evaluator	57
History/rationale of internal evaluation capacity development at org/in industry	80
Having an internal evaluator/liaison made working with external evaluators easier/less intimidating	62
Internal evaluators facilitate external evaluators working with staff	86
Work of internal audience researchers	88
Internal staff with evaluation experience	79
Evaluation efforts elsewhere in organization	59
Development/marketing/other departments asking for/driving evaluation efforts	53
Work with external evaluators	
Work with external contract evaluators	
Work with contract evaluators	5
Contract evaluators primarily for grants/special projects	6
Work with external evaluators limited to director or small number of staff*	58
Contract evaluators work broadly with staff and community	69
Frequency of work with specific evaluators	7
Work with external audience researchers (including market research)	8
Audience research evaluation conducted to meet tax/government requirements	71
Work with university partnerships	50
University partnerships often with students, treated as one-offs	60
How has working with evaluators changed views in evaluation?	11
Haven't worked with evaluators enough for there to be influence	12
Evaluation Culture	
What is an evaluation culture?	
Respondent ideas about evaluation culture	13
Study definition of evaluation culture	14
Use evaluation specifically to improve programs	98
How do you improve your program holistically	63
How does evaluation work have impact on broader industry?	99

Table 23 (continued)

Categories and Codes	Codes
Respondent judgement of department/org evaluation culture	16
Department/staff value evaluation	17
Evaluation culture/value/use can live in pockets w/in department/org	15
Struggle to make time/find resources	18
Lack skills/facility to incorporate evaluation into processes	19
Challenges in utilizing results of evaluations	89
Balance of desire to do more evaluation and need to finish projects (exhibits)	91
How does evaluation culture differ in dept vs. rest of org?	20
Evaluation culture/knowledge/value/use higher in programming/edu department	64
Disagreement/differences in metrics	22
Leadership views of evaluation	
Org leadership value evaluation	21
Leadership support is implied or tacit	75
Leadership not supportive of evaluation efforts	94
Review of survey scores	23
Communication scale	24
Systems and structures scale	25
Would staff score the survey differently?	26
Survey scores by staff would likely be very similar	27
Survey scores by staff would likely be different	28
Staff that are more involved would score higher (more familiar) and vice versa	29
Staff that are more involved would score lower (more critical) and vice versa	30
Staff that don't know the history of evaluation at org might score lower because they don't know how far the org has progressed	82
Evaluative Thinking	
What is evaluative thinking?	
Respondent ideas about evaluative thinking	31
Study definition of evaluative thinking	32
Evaluative thinking as a social process	33
How does evaluative thinking show up in your/department's work?	34
Reflective practice is/not practiced/examples	35
Assumptions/positionality is/not practiced/examples	36
How DEI influences evaluative thinking/evaluation culture (esp. assumptions/positionality)	49
Systematically collected evidence is/not practiced/examples	37
Impetus for evaluation/data-driven decision making influenced by science identity of org	81
How does evaluative thinking differ in dept vs. rest of org?	38
Professional development conducted/examples	39
Broadscale training/work with evaluators across department led to improved evaluative thinking practices	65
Informal, team-building-style activities (article clubs, etc.)	66
Internal evaluators providing formal/informal professional development	84
Seeking/using grants to support evaluative efforts/training	40
Professional development through action research	83

Table 23 (continued)

Categories and Codes	Codes
Develop evaluation capacity by participating in evaluations	85
Psychological Safety	
What is psychological safety?	
Respondents ideas about psychological safety	41
Maslow's hierarchy invoked in discussing psychological safety	42
Study definition of psychological safety/respondent reaction to scores	43
Relationship between psychological safety and evaluative thinking/evaluation culture	46
Learning is risky	48
Evaluative thinking related to process	47
Factors positively or negatively contributing to psychological safety	
Positive	
Being open to ideas/input part of personal practice/values	55
Encourage trust/safety by helping staff make decision (which involves risk taking)	76
Leadership modeling risk taking/making mistakes	100
Work with evaluators has lessened fear of or apprehension about evaluation/built trust	61
Provides voice for staff	73
Negative	
Team-related	
Staff turnover/org stability influencing psychological safety	44
Team member personalities can contribute to or detract from psychological safety	67
Team is risk-averse (due to previous bad experiences with leadership)	95
Leadership-related	
Safety linked to concern that leadership does/not value education work/staff	87
Staff don't see leadership as open to ideas or feedback	68
Leadership is risk averse	56
Broader/interrelated factors	
Staff accountability as a contribution to psychological safety	92
Siloing at org diminishes psychological safety (or affects other aspects of evaluation culture)	96
COVID influence work/psychological safety	45
Psychological safety in department vs. org	77
Risk more available to education staff because stakes are lower	78
Unsorted/unrelated	
Defining terms/study clarifications	1
Accessibility assessment	72
Co-design/co-creation	90
ROI as part of the evaluation process	93
Are scores influenced because respondents don't know what they don't know?	101
Unrelated/irrelevant	999

Note. Codes reflect interview coding, consistent with Table 13. **Bold denotes top-level theme categories.** *Italics indicate sub-categories (used where necessary)*

In summary, the 100 responding education directors are a reasonably representative sample of U.S.-based, AZA-accredited zoos and aquariums when considering governance, operating budget, and annual attendance. Of responding institutions, 90% indicated at least occasional work with professional evaluators, primarily external evaluators. There were 21% that indicated some kind of internal evaluation staff and 38% that indicated the presence of staff with experience or training comparable to a professional evaluator. Larger institutions were more likely to have invested in evaluation resources. Scores on the phase one survey revealed no significant relationships between work with professional evaluators and the overall evaluation culture of the responding directors' departments. An exploratory factor analysis revealed an emergent construct, psychological safety, that may explain some of the variance in scores. Follow-up interviews with nine case-study institutions provided support for previous research on the benefits of working with evaluators but suggest evaluator influence may be limited by institutional structure or decision-making. Directors expressed resonance with, and interest in, the influence of psychological safety.

Chapter Five: Discussion

To answer this study's primary research questions around whether institutions with internal evaluation staff are associated with stronger evaluation cultures (and whether work with evaluators generally is associated with stronger evaluation cultures) this section synthesizes the results of the survey and follow-up interviews and discusses the emergent construct of psychological safety.

As noted, it is important to remember the context of the study, a global pandemic that had a significant impact on the finances and operations of zoos and aquariums in the United States. As an example, the Seattle Aquarium was closed over 150 days in 2020 forcing reductions or lay-offs for over 40% of staff and a revenue shortfall of $14 million on a $22 million budget. AZA accreditation activities were also suspended. While circumstances varied throughout the country, this kind of impact was typical and likely weighed on the minds of respondents despite instructions to consider pre-COVID conditions. Several said as much during the follow-up interviews:

> We were in [the storming stage of team development] and we never got past that one . . . [the pandemic] unsettled things to the extent that everyone was sort of dealing with the crisis . . . more than half of my team has been in their position . . . for less than three months before the pandemic hit and everything changed. (Institution A: Weak)

> [T]here is still that recognition that things just aren't the way that they were pre-COVID. I have fewer people my department, granted we haven't made nearly the kinds of cuts I know that a lot of other people have, but they've still hurt and we’re still doing more with less . . . I would love to come back to us a year from now, and have the same conversation. Even if nothing has changed. Where were we then versus where are we now, because we're still struggling, and there's no end in sight, which doesn't help either. (Institution C: Weak)

Work with Professional Evaluators

The survey results suggest working with external evaluators in one form or another is common among institutions with 90% indicating some kind of work with external evaluation or research staff. Often, these external evaluators were contract evaluators (mentioned by 70% of respondents) associated with a discreet grant-funded project or exhibit.

> We have a lot of capital projects going on here and as part of those capital projects, we build in evaluation, both front end evaluation and . . . we really go through the whole process. We've done it with every major project we've done here. (Institution D: Moderate)

> It's not something that happens often, but it has happened . . . normally associated with special projects . . . (Institution A: Weak)

> [I]n every single case we've had paid external evaluator, it's always been for a grant. (Institution C: Weak)

University partnerships were also common (mentioned by 67%). These can vary from students conducting projects for their own learning, to students/professors working with staff to understand or evaluate institutional programs, to cooperative research projects. At least one case study institution mentioned the tendency of university partnerships to be treated as a string of single projects, rather than a more comprehensive or ongoing look at the institution's programs.

> [E]ach [university partnership project] has been very different, based on what the professor wants the students to achieve as well as how much time they have to contribute . . . they're very one-off in the sense that none of them are connected to any other ones. In fact, the only one that even was remotely long term was [associated with a particular professor] because there were a couple times where I was able to work with the same students over multiple semesters, but in most cases, it's a one semester deal. They've got things they want to achieve. We've got programs that we're trying to learn about, so we find the right fit and boom, they come in and do something, we learn from it and we move on . . . versus building a formal relationship with [the university partner] to evaluate the things we want to achieve long term. (Institution C: Weak)

Other follow-up interviews also mentioned the episodic nature of university partnerships, but expressed more appreciation for the depth and impact of those engagements, with the partnership bringing expertise and capacity that the institution may not have been able to provide.

> It was a yearlong process with community partners. [We worked with the university and their extension service], who do a lot of work with, really elevating and training small nonprofits and people who are interested in

> community leadership, and really making a difference for their communities. That was such a career highlight . . . .and that's led to us doing more co-creation with our community members around programs. (Institution E: Strong)

> [Our work with our university partner is] a collaborative, evidence-based learning network for improving environmental education distance learning. They are looking at [our] virtual programs, which obviously is all we're doing right now. (Institution G: Weak)

> We work with [a local university partner] on a couple of different things. One of them was a partnership between us, [the university and the local children's hospital] looking at accessibility. We have a professor of their behavioral analysis area and [their] grad students that do regular training for us. As part of it, they came and actually walked around with families to evaluate our facility as far as looking at our accessibility in different ways. They did an initial study to roll out the program, and then we've done a few follow ups, every time we do a training, [to evaluate the changes we've made]. (Institution D: Moderate)

Work with external audience researchers was mentioned by 56% of respondents. There are several consultancies that work with zoos, aquariums, museums, and other cultural institutions to help them understand their audiences, both in the context of visitation and in relation to the surrounding community (market). These efforts are often around increasing attendance and visitor experience satisfaction. The Morey Group is one such market research firm with which the Seattle Aquarium has experience and who was mentioned by several larger case study institutions. Working with these firms requires a significant investment of resources and is less likely among smaller zoos and aquariums. At least one case study institution also mentioned audience polling and assessment required by and paid for a local tax measure.

Internal evaluation resources. A fifth of responding institutions (21%) indicated some kind of professional evaluation capacity on staff. That included 15% where those resources represented one or more program evaluators. Most program evaluators were situated in programming departments, but internal audience researchers were just as commonly located in marketing or communications departments. Nearly 40% of all respondents mentioned the presence of internal, non-evaluation staff with experience or education comparable to professional evaluators (by their estimation).

Only one aspect of institutional demographics was connected to work with professional evaluators and that was institutional size. Both operational budget and annual attendance showed a positive association with the presence of internal evaluation staff (p = .00 and p = .01, respectively, with Cohen's d effect sizes of 0.79 and 0.87). Additionally, scores on one item in the evaluation dimension ("The integration of evaluation activities into our department's work has enhanced (or would enhance) the quality of decision-making.") showed significant differences between institutions when sorted by operational budget (p < .01) with Bonferroni post hoc tests showing differences between the smallest institutional budget category (annual budgets less than $2 million) and both the medium-large and large institutional categories. While all of the evaluation items were written to allow for either current or potential evaluation activities, smaller institutions may not have seen their internal and informal evaluative processes as *evaluation activities* and therefore been less likely to acknowledge a benefit. Overall, this correlation between institutional size and investment in evaluation should be expected. As the three barriers to conducting evaluation most commonly cited in the literature are money, expertise, and time (Clavijo et al., 2005; Khalil & Ardoin, 2011; Luebke & Grajal, 2011; Ogden & Heimlich, 2009; Roe et al., 2014), it stands to reason that larger institutions with larger budgets may be in a better position to invest in evaluation. These larger budgets may also be able to afford to pay competitively for more experienced or educated staff in leadership roles that may, in turn, be more likely to value and advocate for investment in evaluation resources.

The experience, education, or perspective of the education/engagement director is a factor not explored explicitly in this study that might also influence institutional decisions to direct funds towards evaluation resources. All nine directors in the case study interviews expressed strong value for evaluation. If that were taken at face value, one might argue that valuing evaluation or evaluative thinking is not necessarily related to investment in evaluation resources (or an institution's evaluation culture), but it is very possible that this response was influenced by the self-selected nature of participation in this study as well as both response and social desirability biases. As noted, my profile within the zoo and aquarium community and/or familiarity with survey and case-study interview respondents could be argued to mitigate or exacerbate these effects. None of the nine directors noted direct experience as evaluators or professional training. Several mentioned some exposure to evaluation ideas and practices in graduate school course work and/or through sessions or workshops at conferences. Most characterized their experience/understanding as acquired through informal activities (developing their own evaluations, working with external evaluators). At least one noted being influenced by working with previous leaders who valued and supported evaluation. This cultural transmission of values is another interesting area to explore further.

Relationship Between Evaluators and Evaluation Culture

This study hypothesized that institutions with internal evaluation capacity would be positively associated with the strongest evaluation cultures. Working with evaluators has long been valued for the process benefits of the experience (Hargreaves & Podems, 2012; Patton, 1998; Preskill & Zuckerman, 2003). Working with evaluators can build the evaluation skills of staff and reduce anxieties associated with evaluation (Labin et al., 2012; Monroe et al., 2005; Suárez-Herrera et al., 2009; Volkov, 2011), increase demand for evaluation (Beere, 2005), and even serve as a *catalyst for change* in creating a more evaluative culture (García-Iriarte et al., 2011). Some of these benefits could be accrued through work with external evaluators, *if* that work was frequent enough, but it is unlikely that work with external professionals could be as frequent or consistent as work with in-house staff. This was reinforced by the findings of the survey in which work with external program evaluators on a monthly or weekly basis was only mentioned four times. Hiring internal evaluation staff is a signal of priority by institutional leadership. Even if it is a strong evaluation culture that leads to the hiring of internal staff, rather than work with internal evaluators leading to a stronger evaluation culture, a positive association should be expected.

Evaluation culture as viewed by respondents. Evaluation culture was assessed by the modified Readiness for Learning and Evaluation (ROLE) instrument in phase one of the study. The modified instrument consisted of 50 items, each on a 0-100 scale (strongly disagree to strongly agree). The instrument contained six subscales or *dimensions* (organizational culture, leadership, systems and structures, communication, teams, and evaluation), with 2-20 items each. Dimensional means were calculated, and an *overall evaluation culture score* was calculated by averaging the six dimensional means. The score represents a director's subjective assessment of the culture of their workgroup and is a product of their personal experience within both the organizational and broader cultural context.

Dimensional means ranged from 60.79 (communication) to 80.64 (leadership) (see Table 17). From one perspective, communication is often identified as a challenge for, and important driver of, employee engagement (Kular, Gatenby, Rees, Soane, & Truss, 2008; Zajkowska, 2012), so seeing a lower score might be expected. On the other hand, there were only two items in this scale with the mean on one item ("There are adequate records of past change efforts and what happened as a result.") dragging down the mean of the other item ("Information is gathered from guests, program participants, and/or other stakeholders during department activities to gauge how well we're doing."), which was closer to the other dimensional means (*M* = 49.36 versus *M* = 72.22). A number of communication-related questions were eliminated from the original instrument as it was trimmed to increase focus and reduce the time required for completion, which resulted in a two-item scale vulnerable to score effects. A higher leadership score might also be expected as respondents are all part of *leadership.* Higher scores might be a product of possessing more complete information on the opinions/attitudes of department and organizational leadership, or they may have been influenced by social desirability bias. The overall evaluation culture mean of 74.79 (on a 0-100 scale) suggests that respondents generally agreed that their workgroup possessed a reasonably strong evaluation culture.

The relationship between work with professional evaluators and an organization's evaluation culture was explored from various perspectives. No relationships were found between work with internal evaluators and evaluation culture scores, either when explored as a group (program evaluators *and* audience/social science researchers) or separately. A similar pattern (or lack thereof) was found when exploring the relationship between evaluation culture scores and external evaluators, with the presence/absence of trained internal (non-evaluator) staff, and in all combinations of the three.

There are several potential explanations for these findings. Following Ockham's Razor, the most straightforward explanation could be that there is no relationship. A strong evaluation culture in a workgroup or institution may emerge (or not) independently of the presence of internal evaluation staff. It may also be agnostic to work with external evaluators. This could be viewed as a positive finding, especially for smaller institutions. While time, expertise, and funding may be barriers to conducting formal evaluations, they may not prevent a workgroup from developing a consistent practice of evaluative thinking, with formal evaluations conducted periodically on a project basis. That mindset is demonstrated in this quote from an interview with an institution that indicated no work with professional evaluators, but also scored itself a strong evaluation culture based on its responses to the survey (overall evaluation culture score: 82.17):

> [It is] always, how can I improve it? . . . How can I make it better? . . . Is it helping? . . . to me, evaluative thinking should be on focused what our objectives are. And sometimes we get off into the weeds and get a little bit more hyper-critical than we should be, but overall, I think, if you're a good educator, you're evaluating the whole time, even when you're teaching your way, because you're watching your audience, and you're seeing what they respond

> to, and what they don't. So, I think educators just tend to be natural self-evaluators and have evaluative thinking. (Institution I: Strong)

Another possible explanation could be that the instrument or study design was not effective at detecting the relationship. The original ROLE instrument was designed for general use in organizational settings. Zoos and aquariums are organizations, but with features that make them distinct. Many/most operate with a tension between business and mission objectives. Staff performing organizational activities typically associated with *business* objectives (finance, marketing, communications, human resources), may not have science or conservation backgrounds and sometimes feel disconnected from the mission activities of the institution. Conversely, staff associated with *mission-related* activities (education/engagement, animal care, conservation, research) may view business objectives (increasing revenue, attendance) as merely a means to an end, even distasteful. It is not clear if these organizational dynamics would compromise the appropriateness of the instrument for zoo and aquarium settings, but any time an instrument is applied to a new audience, there is potential for variance. Considering respondents came primarily from the programming departments of each institution, these tensions may not have come into play.

The instrument was also modified in several ways. Language changes likely increased fit for the sample audience, increasing clarity of the intended context for each item, but a number of questions were removed to shorten the instrument in the interest of increasing completion rates. Removing these questions could have compromised the effectiveness of individual subscales (e.g., communication), or the instrument overall. Two questions were also added to the evaluation dimension to better address study constructs of evaluation culture and evaluative thinking. While not part of the original tool, they were included in validity reviews and were not flagged by reviewers. The corrected item-total correlation for item eight (E8) on the evaluation subscale was .42, while for item seven (E7) it was .31. However, removing E7 would not have resulted in an improvement of the alpha for the dimension, so it was not excluded. One other concern was expressed by several respondents around the appropriateness of the survey for very small departments. Some smaller institutions operate with as few as one programming staff person (one education staff person is a minimum requirement of AZA accreditation). For programming departments with 1-2 staff, questions about how department management view evaluation versus staff might not be very meaningful. Some small institutions (reported through personal communications) opted to fill out the survey with their full organization in mind. This is a different framing context than was intended and could have affected how the smallest institutions reported their scores.

While the case for alignment with the study constructs was made in Chapter Three and generally supported by reviewers during content validation, there were some concerns that the items did not possess enough nuance to adequately interrogate the construct of evaluative thinking. This was addressed in the study design through the inclusion of the follow-up case-study interviews, but it could have affected evaluation culture scores. One reviewer recommended reverse wording for some/most items, but discussion with other advisors elevated concerns about trade-offs between the benefits of reverse wording in addressing acquiescence or confirmation biases and confusion created by inconsistency. Additionally, even if the modified survey instrument was perfectly aligned with study constructs, it is not possible to include every factor that might influence or serve as antecedents to those constructs. An example already noted is the experience, education, and attitudes of the director. While leadership featured prominently in the literature around evaluation culture, the conversation centered on leadership support, provision of resources, etc. How might the personal motivations to incorporate evaluative thinking have influenced their support or strategy? While many may express personal interest and support for evaluation, do they have the skill set to identify and/or prioritize professional development needs? Do they have the skillset to manage change in a team or organization?

Finally, as a self-reported and voluntary survey, response and social desirability biases could have affected results. Study definitions were removed on advice of reviewers in an attempt to reduce satisficing, but the short purpose statement at the beginning of the survey and in the invitation (deemed necessary to secure participation) could have cued participants to the *right* answers. These biases are could have elevated individual dimension or overall scores, but they are less likely to have encouraged embellishment of an institution's work with professional evaluators. If these biases artificially inflated scores or reduced score variance, it may have been more difficult to detect an effect. As one interviewee put it, respondents may also be limited by their own experience, "maybe those of us that don't have professional evaluators just maybe don't know what we don't know?" (Institution I: Strong). It is possible that any respondent may not have sufficient variability in their own professional experience to judge whether their work group is high or low performing in a particular aspect of evaluative thinking or culture (as characterized by individual survey items). Some with less experience or lower standards may score their departments more highly while those with more experience or who may be more critical, may score their workgroups lower. This is a challenge with any self-reported questionnaire that asks respondents to make a subjective judgement.

It is also possible no relationship was observed because the conditions were not favorable at enough institutions for evaluators, and internal evaluators in particular, to have an influence. Volkov (2011) outlined eight interrelated roles for internal evaluation staff to support a developing evaluation culture. They include: (a) change agent, (b) educator about evaluation, (c) evaluation capacity building (ECB) practitioner, (d) support for management decisions, (e) consultant, (f) researcher, (g) advocate, and (h) promoter of organizational learning. While not covered in the survey, the three institutions in the case study interviews of phase two shared examples of internal (and external) evaluators filling many of these roles. However, opportunity for staff to work with evaluators was limited by organizational placement and decisions made around how to manage these resources. Several examples follow.

Institution F is a large, well-resourced organization with internal audience research capacity and at least some trained internal (non-evaluator) staff. They work with external evaluators on a project basis. The staff from the programming department work with the institution's audience researcher (located in their marketing department) as a consultant on elements around the promotion and popularity of interpretive programming among members and guests.

> [They are] in charge of the exit surveys . . . [they will] do testing on names and even descriptions to see what people are mostly interested in . . . [they will] do focus group testing for memberships . . . and then when [they have] worked with us in programming and it has been a few times over the last 10 years . . .

> it's not primarily what [they do] so . . . most recently, it was something we wanted to know from members . . . [they] included a couple programming questions, [they have] done that in the past . . . and something that [they] did specifically for us was around conservation asks, and what people already felt like they did, and then what they would be willing to do and that was done, oh, eight, nine years ago maybe. (Institution F: Moderate)

Institution F also described similar, limited consultive work with trained internal (non-evaluator) staff in their school programs department. The education director from Institution A, a medium-large public institution with some exposure to external audience researchers contracted by the organization, also described occasional opportunities to utilize the resource.

> I mean, when needed. So it's run through the marketing division of [the organization], so they're the ones who are coordinating that. We will occasionally go in and add questions [to visitor surveys] when we have some specific things we want to find out. We, a couple years ago, wanted to know a little bit more about the perception of animal care, so we added some questions related to that. Programming kinds of questions, when programs change . . . but it's kind of, you know as needed. (Institution A: Weak)

Their work with these resources gave them a chance to interrogate their work and answer questions with data. There is an opportunity to test assumptions (e.g., what member programs members are interested to see). There is some tacit support from management, so much as there is support for the audience researcher’s time for this work, but their work with these resources was likely too infrequent to support or influence an everyday culture of evaluative thinking.

Institution F's survey responses placed them in the middle (moderate) tier when looking at overall evaluation culture scores (79.59 out of 100). Institution A was in the lowest (weak) tier with a score of 55.58.

Institution C is a large, well-resourced organization that indicated one FTE of internal evaluation staffing and regular work with external evaluators. Their evaluator is new in the role and is helping them better plan and organize their evaluation efforts. They (and trained, non-evaluator staff before them) have provided benefit as consultants, supporters of management decisions, and especially as advocates of evaluation.

> [T]here had been [among staff] more of a fear of evaluation as [something that would] show us we're doing it wrong . . . almost a sense of, we have this person coming in and judging us . . . and I wouldn't say I ever felt that flippantly about it, but even when I first started working with evaluators I kind of had this perception or this concern of, you know what, if we've got this several hundred thousand dollar [grant], we get to the end of it and realize that we didn't achieve the outcomes even remotely that we set out to achieve . . . how's it going to look? And so I think, over time, we've all learned to appreciate it and realize that this process is about improvement and it's *also* about an obligation to our funders, in the sense to that we want to be able to demonstrate that we're fiscally responsible in the sense that we do the work that we say we're going to do. And if we don't achieve the outcomes that we set out to achieve, that we've got a plan for how we would approach things differently. Going forward, [in our department] there's a much a much stronger appreciation for working with

> external evaluators than there had been maybe 10-15 years ago [that comes from a] combination of [working with evaluators and the evolution of our field, but more heavily weighted on working with evaluators. (Institution C: Weak)

Institution C felt working with evaluators (internal and external) has demystified the process and reduced anxiety around evaluation for their department staff, but other factors have limited the influence of evaluators on the development of an evaluation culture. Institution C's responses to the survey placed them in the bottom tier of zoos and aquariums in the study in regard to overall evaluation culture score (69.69). Their internal evaluation staff person was hired only recently, went out soon after on an extended leave, and then the pandemic struck. Their opportunity to have an influence on Institution C's culture has been limited.

Institution E is another large, well-resourced organization with an independent evaluation and research department, multiple trained (non-evaluator) staff, and frequent work with external evaluators. Institution E scored near the top of the top tier of the sample with an overall evaluation culture score of 89.59 based on their responses to the survey. The evaluation department at Institution E works in all of the eight of the roles outlined by Volkov (2011). The development of this evaluation capacity was driven by a long-tenured education director.

> So ever since I've been working in the zoo's education department [evaluation has] been something that's really important to me . . . We say we're making changes, blah blah blah, but how do we know? . . . I wanted to really increase the quality and the impact of our programs and knew we had to kind of move out of the just, well, this is what I *think* . . . (Institution E: Strong)

Institution E’s education director worked with internal and external peers locally to leverage each other’s resources to do more evaluation. Over time, with additional institutional and grant support, they were able to build more internal capacity, which in turn, had an influence on the culture of the organization.

> [B]eing around it, sharing studies, sharing results and then there's people saying, hey, I want to evaluate this, can you help me? So that's how the culture got built and it's an, and as a site, as an institution that's based on science, you know, we've gotten to where stuff that used to be opinion battles, not just in [our department] but the whole institution, you know, or who's been here the longest or who talks the loudest or who can bully everybody else. It's turned to be like, well, what does the data show us? Let's ask some questions. Let's find out what do we know . . . Well, it's been a long path, but it's one of those, like, if you set your mind to it, and you play the long game, you'll get there, and where we are now where I've been wanting us to be for a long time, but you just keep working it . . . (Institution E: Strong)

Institution E is a model example where a leader with vision and value for the role of evaluation created capacity and culture around reflective practice, social learning, and data-driven decision-making. Their evaluation and research department was (and is) able to provide time, expertise and resources to other staff on a weekly or even daily basis.

Institution D was also able to provide consistent and predictable availability of evaluation capacity through regular engagement with external evaluators. Institution D is a medium-large organization with no internal evaluation staff, but that regularly works with external evaluators through their capital projects.

> We have a lot of capital projects going on here. And as part of those capital projects, we build in evaluation, both front-end evaluation and [summative] . . . When I started here in 2017, one of the first things I did is bring on an evaluator, as well as interpretive design consultant, and they worked together on the interpretive planning for the exhibit, so they kind of brought us through that process. They worked in the community. They worked with staff; they worked with volunteers, and they have helped us. They are still actually helping us in the process of [new exhibit], and we will probably shortly be bringing somebody on for [different new exhibit], which is our next capital project. (Institution D: Moderate)

Although not part of organizational or department staff, the regularity of capital projects and the consistency of the policy of involving evaluators from the beginning of the projects, created regular work and exposure to evaluation practices. The director from Institution D also talked about how work with these evaluators helped reduce anxiety or misconceptions about evaluation among staff and helped build demand and interest in data-driven planning.

> Outside of those capital projects, when we get a new species into our collection [for example] . . . [department staff] write all that content, does the research behind what's going to go on those signs . . . and I do think that that same level of, well, if we're going to say this, we should know where it's coming from, and we should know who the audiences are, and we should know who we could ask

> within our community, who could we approach? All of those questions, I think, are brought to the forefront because of the work we did on [our last capital project]. What is it, a rising tide raises all ships? (Institution D: Moderate)

The responses from Institution D on the phase one survey placed them right at the mean of overall evaluation scores (76.79). This score is not in the top (strongest) tier, but the mean does reflect agreement that a healthy evaluation culture exists, and it is difficult to say what their organization's culture would look like without this regular input from external evaluators.

A common theme that emerged from follow-up interviews arose around *who* is exposed to an institution's work with evaluators. Institutions E and D are examples where evaluators have broad, regular contact with multiple staff within the institutions' programming departments. In other cases, work with external evaluators was limited to a single point of process contact with some additional contact through reporting. The education director from Institution C shared that before they added internal evaluation staff, project work with external evaluators was limited to themselves and/or the lead staff person of the program/area being evaluated. This was exacerbated by the limited resources (time) for evaluation built into grant budgets. Similarly, Institution F shared that exposure to the evaluation process was limited to a single or several lead program staff associated with the evaluated program. At smaller institutions like Institutions B, G, and H, this was also the case, sometimes out of necessity due to the small number of staff in the department. Limiting exposure to evaluation processes to a few staff or only occasional circumstances (as the example of Institutions F and A's limited utilization of audience research resources) would necessarily dampen any potential impact that exposure might have on a department or organization's evaluation culture. The education director at Institution A said as much when asked how work with professional evaluators did or did not affect the view or value of evaluation in the organization, “we don't work with them enough in order for it to have access, so we’re just not getting enough time with them to feel like, to have a bigger influence.” (Institution A: Weak) These limitations may be the product of simple inertia of practice (“the way we’ve always done things”) or may be the result of structural and managerial priority decisions around resource use.

While the phase one survey results did not indicate a relationship between work with professional evaluators and an organization’s self-reported evaluation culture, several case study institutions shared the positive influence of that work (in addition to the examples above). Institutions D and C felt working with external evaluators helped build evaluation knowledge and skills with Institution C feeling it complemented and bolstered work done in professional development by the department. The staff at Institution E (with an independent evaluation and research department) benefit from both professional development provided by internal evaluation staff and action research opportunities either created by the evaluation staff or facilitated by the staff through partners. Lastly, Institution D shared an interesting way that evaluation efforts that are well integrated into everyday practices across the department can potentially enhance employee engagement, “I think it's a way to give a voice to teams that often feel like they don't have a voice, or that it's not always reflected in what is seen in the [institution].” (Institution D: Moderate)

Psychological Safety as an Emergent Construct

Although there were no statistically significant relationships between evaluation culture scores and any variables associated with institutional demographics or work with professional evaluators there was still variability in scores. An exploratory factor analysis revealed four factors with some explanatory potential. Factor one included 19 items across all six original dimensions which aligns with many of the component constructs of the study's definition of evaluative thinking, with the exception of the idea of interrogating assumptions and the addition of leadership support . . . itself a key component of the study definition of evaluation culture. (see Table 24).

Factor four consisted of seven items from three of the original subscales, primarily the evaluation dimension, that also align well with learning, growth, and evaluation. The following items were included in this new evaluation/growth dimension (with original subscales in parentheses):

- Department employees are recognized or rewarded for learning new knowledge and skills. (systems and structures)
- Department employees are recognized or rewarded for experimenting with new ideas. (systems and structures)
- Teams and work groups in the department are encouraged to learn from each other and to share their learning with others. (teams)
- The integration of evaluation activities into our department's work has enhanced (or would enhance) the quality of decision-making. (evaluation)
- Evaluation helps (or would help) the department provide better programs, processes, products and/or services. (evaluation)
- There would be support among department employees if we tried to do more (or any) evaluation work. (evaluation)
- Doing (more) evaluation would make it easier to convince department and organizational leadership of needed changes. (evaluation)

Table 24

Items Aligned on New Evaluative Thinking Dimension

Items	Components of Evaluative Thinking
• Managers and supervisors in the department take on the role of coaching, mentoring and facilitating employees' learning. • The current reward or appraisal system in the department recognizes, in some way, team learning and performance. • Our department currently operates via (or is transitioning towards) a team-based structure. • Department employees are provided adequate training on how to work as a team member. • Team meetings in the department address both team processes and work content.	Social learning
• Department employees continuously look for ways to improve processes, products and/or services. • Department employees are provided opportunities to think about and reflect on their work. • When trying to solve problems, department employees use a process of working through the problem before identifying solutions. • Department employees continuously ask themselves how they're doing, what they can do better, and what is working.	Reflective practice
• Department employees use data/information to inform their decision-making. • Managers and supervisors in the department use data/information to inform their decision-making. • Information is gathered from clients, customers, suppliers or other stakeholders during department activities to gauge how well we're doing. • There are adequate records of past change efforts and what happened as a result. • There are evaluation processes in place that enable department employees to review how well changes we make are working. • Data are routinely collected during department activities to inform evaluation efforts.	Data-driven decision making
• Managers and supervisors in the department view individuals' capacity to learn as among the organization's greatest resources. • Managers and supervisors in the department help employees understand the value of experimentation and the learning that can result from such endeavors. • Managers and supervisors in the department provide the necessary time and support for systemic, long-term change. • Department employees are recognized or rewarded for helping solve organizational problems.	Leadership support

Note. Study definition of evaluative thinking: *a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making*.

Factors two and three aligned with a new construct not uncovered in the original literature review, this idea of *psychological safety*. Psychological safety is a term coined in the 1960s and brought back to the conversation around workplace learning and performance through Kahn (1990) and Edmondson (1999). The definition most commonly cited in the literature is from Edmondson (1999): "a shared belief held by members of a team that the team is safe for interpersonal risk-taking" (p. 350).

Learning is an inherently risky act. Asking for help or admitting errors are necessary for learning and growth but can result in a loss of *face* (Brown, 1990). Employees have a tendency to hide errors or knowledge gaps to mitigate this threat, but this inhibits team and organizational learning (Argyris in Edmondson, 1999). When teams possess a strong sense of psychological safety, they are confident they can reveal these errors and gaps without concern for embarrassment, rejection, or punishment from the team or its leader (Edmondson, 1999). This was consistent with Kahn (1990) who suggested employees should be able to "show and employ one's self, without fear of negative consequences to self-image, status, or career" (p. 790). Overcoming this learning anxiety comes from seeing positive results from this kind of risk-taking (Schein, 1993).

Edmondson (1999) distinguishes psychological safety from trust, defining trust as, "the expectation that others' future actions will be favorable to one's interests, such that one is willing to be vulnerable to those actions" (p.354). Psychological safety goes beyond trust, requiring team members to also feel mutual respect and care for one another as individuals. Newman, Donohue, and Eva (2017) further distinguish that trust is something that individuals feel for each other, while psychological safety is a shared, group construct. This is consistent with the way Edmondson and others have discussed psychological safety (at the group or organizational level), but it must be measured at the individual level therefore necessitating an individual-level aspect of the construct.

Newman et al. (2017) note that Kahn's definition (Kahn, 1990) was more about how the individual felt than the group, but that Edmondson's definition has had more traction precisely because of its group-level approach. As it happens, several reviews have noted that associations between common antecedents and outcomes of psychological safety are consistent across individual, group, and organizational-level analyses (Edmondson & Lei, 2014; Frazier, Fainshmidt, Klinger, Pezeshkan, & Vracheva, 2017; Newman et al., 2017). However, several also suggest that psychological safety is most salient at the group level and that significant variance can be see across teams within the same organization.

Team psychological safety has a positive relationship with group learning, defined in Edmondson (1999) as "an ongoing **process of reflection** and action, characterized by **asking questions**, seeking feedback, **experimenting**, **reflecting** on results, and **discussing errors or unexpected outcomes of actions** . . . [where] team members must test **assumptions** and discuss differences of opinions openly. . ." (p.353). The bolded areas indicate overlap between this conceptualization of team learning the study definition of evaluative thinking:

> *Evaluative thinking is a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making.*

One key difference between the two definitions is evaluative thinking's link to decision-making. Part of the risk of learning activity is that it can consume time and resources without the guarantee of results (Edmondson, 1999). Decision making is core to the construct of evaluative thinking just as the idea of judgement (of merit or worth) is core to evaluation (Scriven, 1991).

Learning-associated behavioral outcomes linked to psychological safety include seeking feedback, sharing information, asking for help, talking about errors, and experimenting (Edmondson, 1999; Edmondson & Lei, 2014; Newman et al., 2017). Psychological safety is also linked to other performance and organizational growth-oriented outcomes and behaviors like employee engagement, satisfaction, voice, communication, and knowledge sharing (Edmondson & Lei, 2014; Frazier et al., 2017; Newman et al., 2017).

In most models, psychological safety is seen as a mediator between certain antecedents and learning or performance outcomes. These antecedents include leader style and behavior, supportive team and organizational structures (including access to resources), individual mindset (especially learning orientations and proactivity), status, and high-quality inter-personal relationships (Carmeli, Brueller, & Dutton, 2009; Edmondson & Lei, 2014; Frazier et al., 2017; Newman et al., 2017). While all these antecedents themselves contribute to learning and performance, a meta-analysis by Frazier et al. (2017) found evidence across studies that psychological safety provided additional explanatory power over unmediated antecedents.

A seven-question item set from Edmondson (1999) has been used or modified frequently throughout the literature on psychological safety since its publication (Edmondson & Lei, 2014; Frazier et al., 2017; Newman et al., 2017). These items align with factors two and three from the exploratory factor analysis and contain a mix of items that share characteristics with Edmondson's (and others) definition of psychological safety. Factor two includes items related to the leadership antecedent of psychological safety and factor three consists of items related to overall team psychological safety. See Table 25 for a comparison between Edmondson's original seven-question item set and the items included in factors two and three. Colored shading identifies thematic similarities between the item sets.

Treating the four factors as new subscales for the instrument accounts for 42 of the original 50 items. The following eight items (with their original subscales in parentheses) did not load (< .36) on any of the four factors:

- Department employees ask each other for information about work issues and activities. (organizational culture)
- Department employees often talk about the pressing work issues we're facing. (organizational culture)
- In meetings, department employees are encouraged to discuss the values and beliefs that underlie their opinions. (organizational culture)
- Managers and supervisors in the department believe that success depends upon learning from daily practices. (leadership)
- Managers and supervisors in the department support the sharing of knowledge and skills among employees. (leadership)
- There is little bureaucratic red tape when trying to do something new or different in the department. (systems and structures)
- There are few boundaries between department units or working groups that keep employees from working together. (systems and structures)
- When the department engages in evaluation activities, the goal is to improve programs. (evaluation)

All four new dimensions demonstrated strong internal consistency (with alphas all above .77). Since the new dimensions also contain all the key constructs associated with the study definition of an evaluation culture—*where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions . . . with the mandate and support of organizational leadership*—calculating means for each new dimension and an average of the four means would create a new version of an overall evaluation culture score. The new overall evaluation culture score also showed strong internal consistency ($\alpha = .80$), had similar measures of central tendency, and a similarly shaped, if slightly more leptokurtic distribution. Institutions that scored highly on the original measure also scored highly on the new measure; institutions that scored low, also scored low. There was some variability at the transition points of the three scoring tiers (strong, moderate, low).

Also consistent with the original measure, there were no significant relationships identified when looking at the new overall evaluation culture score and either institutional demographics. There were, however, differences noted in some measures of psychological safety between the medium-sized and large institutions. In these cases, the medium-sized institutions showed higher mean scores on measured of psychological safety than their larger peers. Institutions in the small category also had lower scores, but the differences were not seen as significant. What is it about the middle range of institutions that led to greater reported feelings of psychological safety? Larger institutions have more structure and hierarchy, typically have greater communication challenges. With more staff, managers may have more direct reports and personal contact and coaching may be limited. Any of these factors may lead to challenges in the key areas outlined by Edmondson and others in their definitions of psychological safety (including, leadership openness to input, risk-taking, collaborative spirit, *inter alia*). Smaller institutions may have been facing greater instability during the pandemic.

Table 25

Similarities Between Edmondson (1999) and Factors Two and Three

Items from Edmondson (1999)	Items from factor two: Leadership-related psychological safety	Items from factor three: Team-related psychological safety
People on this team sometimes reject others for being different.	Department employees are encouraged to offer dissenting opinions and alternative viewpoints.	Department employees are not afraid to share their opinions even if those opinions are different from the majority.
It is safe to take a risk on this team.	Department employees are encouraged to take the lead in initiating change or in trying to do something different.[a]	Department employees feel safe explaining to others why they think or feel the way they do about an issue.
Members of this team are able to bring up problems and tough issues.	Asking questions and raising issues about work with department leaders is encouraged.	Department employees respect each other's perspectives.
It is difficult to ask other members of this team for help.	Managers and supervisors in the department are open to negative feedback from employees.	Department employees tend to work collaboratively with each other.
Working with members of this team, my unique skills and talents are valued and utilized	Managers and supervisors in the department make decisions after considering the input of those affected.	Team meetings in the department strive to include everyone's opinion.
No one on this team would deliberately act in a way that undermines my efforts.	Employees are recognized or rewarded for helping each other learn.	Department employees operate from a spirit of cooperation, rather than competition.
If you make a mistake on this team, it is often held against you.	Mistakes made by department employees are viewed as opportunities for learning.	
	Department employees are confident that mistakes or failures will not affect them negatively.	
	Managers and supervisors in the department model the importance of learning through their own efforts to learn.[a]	
	Managers and supervisors in the department like (or would like) us to evaluate our efforts.	

Note. Colored shading denoted similarities between item themes.

[a] Edmondson (1999) also notes that leaders demonstrating learning behaviors reinforces psychological safety in the team.

There were also some differences noted in team-related psychological safety where in each case the absence of evaluators was associated with higher means. As evaluation can cause anxiety in some staff who may perceive it as threatening, lower team-related psychological safety might be expected where evaluators are present, but this runs counter to opinions expressed by multiple interview participants who felt that working with evaluators helped de-mystify the process and reduce anxiety.

Case study interview participants responded with interest upon raising the question of psychological safety. Although they may not have used the term, the elements of the construct aligned with work they had been doing with their teams.

> It's something that I certainly value. I try and make that the culture that I have within my team of two, but even with our volunteers. I don't want them to feel like they can't share ideas, even though sometimes I feel like that's I'm saying. (Institution B: Strong)

> Yeah, I think this [has] been a big focus for me . . . [O]ne of the things from the beginning, very intentionally, has been about trying to help people to make decisions on our team, like our managers, but also help people all down the line to be able to make decisions. (Institution D: Moderate)

> Yeah, and we don't use those words, but we've spent a lot of time . . . we've worked on the department culture for years, in various ways . . . how we communicate with each other, working on giving feedback to each other . . . both up and down and across and always making sure that staff who don't have official power feel like they have power, and in different aspects of their

> job and what would make them feel more powerful, what would make them have a bigger voice, and be heard and so, while we don't use ‘psychological safety,’ we have been working hard for years on that because it hasn't always been that way. (Institution F: Moderate)

> We've talked a lot about that this year . . . to the point that I about got sick of it because I just, it was a tough year. (Institution H: Moderate)

For others, the idea made them think again about some of the struggles they have faced with their teams, especially in the past year.

> Everyone is new and then we didn't get to do [team-building] in a way that we wanted to do it because, you know, the world didn't let us. Reflecting on that, I think that set us back more than I realized, just in terms of everyone feeling safe and successful and knowing what their role was going to be because we haven't had the time for everyone to learn how to work together and be a part of this team, and so that's why I think that team-related psychological safety score being low (Institution A was in the bottom third of team-related psychological safety scores with a mean of 60.83 on those items) is probably fair and reflective of where we were at and where we probably are at this point. (Institution A: Weak)

Interviewees identified links between psychological safety and evaluative thinking, especially around risk-taking and innovation.

> [I]f they're feeling safe, they're going to be more willing to look at things and give an honest opinion versus what they think people want to hear. (Institution G: Weak)
>
> [E]xactly the bullets that you have up here [covering the key aspects of psychological safety] is how we need to be able to work together. We need to be able to bring ideas to the table, we need to feel okay with making a mistake. It's not the end of the world . . . (Institution H: Moderate)

Though the course of the nine conversations, respondents identified a number of factors that they associated with increasing or decreasing psychological safety at the team or leadership level. Positive influences included being open to ideas from staff [“I feel like a theme for 2020 is like, let's just try stuff.” (Institution D: Moderate)], encouraging trust by empowering staff decision-making, leadership modeling risk-taking, reduced anxiety around evaluation through work with evaluators. Negative influences associated with team-related psychological safety included staff turnover, organizational instability, personality conflicts, and risk aversion associated with previous negative consequences around mistakes. Negative influences associated with leadership related psychological safety mentioned by participants included concerns about leadership support of programming departments, a lack of openness to new ideas from staff, and leaders modeling risk-averse behavior. These influences are sometimes clearest when they change.

> I will say, there were times I felt like I had to put armor on to go to work, you know, and I don't really feel that way anymore . . . I think that does definitely translate down to my staff because, when they're feeling anxious . . . It's like we

> can talk it through, we can bring things to [my supervisor] if we need to, help give us guidance, it's, it's so important. It's everything. (Institution E: Strong)

Siloing within and between departments at the institution, and COVID-related personal and organizational stress were also mentioned.

It's possible that, in the context of a global pandemic, that the items associated with psychological safety might demonstrate more variability and therefore appear to have more of an influence on the variability of scores than they would in pre- or post-pandemic conditions. However, the consistencies between the antecedents of psychological safety (e.g., reflection, inquiry, leadership support) and the study definitions of both evaluative thinking and evaluation culture seem to suggest that the stressors injected into the workplace by the COVID-19 pandemic may have revealed, rather than inserted, this prerequisite of an evaluative or growth mindset, just as the broader cultural pressures of 2020 (climate-induced fires, social justice, and health care) revealed problems already underlying our global systems. As learning and improvement requires being vulnerable and accepting some degree of risk, a work climate of psychological safety is a logical precondition.

While the emergence of psychological safety as an influence on a workgroup or institution's evaluation culture is a potentially interesting finding, it is important to point out that this study was not designed to measure psychological safety as a construct. As Edmondson and Lei (2014) note, where measures are not explicitly designed for and consistent with the construct, there can be validity concerns. Additionally, intentional studies at multiple levels of analysis would be necessary to really explore the role of psychological safety in creating a culture of evaluation within a team. Still, Sanner and Bunderson (2015) note that psychological safety might be particularly important when the work in question is inherently more uncertain or dependent on learning. Considering the context of a global pandemic and widespread industry cutbacks and lay-offs, it might not be surprising that psychological safety emerged as an additional construct, if it was not already present.

Chapter Six: Conclusions, Implications, and Recommendations for Future Research

A survey of 100 education directors at U.S.-based zoos and aquariums were surveyed to examine how a workgroup or organization's evaluation culture varied according to their work with professional evaluators. Evaluation culture was determined through the use of a modified version of the Readiness for Organizational Learning (ROLE) instrument (Preskill & Torres, 2000b) and compared to the presence or absence of internal evaluation staff, work with external evaluators, and/or the presence of trained internal (non-evaluator staff). Follow-up case-study interviews were conducted with nine institutions to learn more details on their work with professional evaluators and its influence on their team's evaluation culture. This chapter discusses conclusions drawn from these studies, implications for professional practice, and recommendations for future research.

Conclusions

When exploring how the evaluation culture of a department or organization varies related to its work with professional evaluators, this study found that work with evaluators is common, but not consistent and can be limited within an organization. Larger organizations (as measured by operating budget or annual attendance) are more likely to have invested in evaluation capacity. However, institutions were no more likely to work with professional evaluators based on their form of governance (non-profit, for-profit, or public).

When reviewing the overall evaluation culture scores compiled from the combined responses across the six subscales of the survey, participating education directors scored their workgroup's evaluation culture relatively high. Survey results from phase one of the study suggested that institutions with internal evaluation capacity were no more likely to be associated with strong evaluation cultures. The inverse was also true, institutions that indicated no work with professional evaluators in any capacity were no more likely to be associated with weak evaluation cultures. Organizations that worked with a combination of internal and external evaluators showed no differences from other institutions. Follow-up interviews with nine case study institutions suggested work with evaluators has been beneficial to their teams, their impact mitigated by limitations in staff exposure, organization structure, and managerial decision making. In addition, psychological safety emerged as an additional antecedent influencing the development of evaluative thinking and an evaluation culture in a team or organization.

Implications for Professional Practice

In addition to the education directors who were the focus of this study, these findings carry implications for a variety of stakeholder audiences including:

- Executives at zoos and aquariums who have the ultimate authority over resource use and organizational culture. This includes both chief executive and operating officers and the heads of other departments. This study was delimited to the education/programming departments of zoos and aquariums, but a growth mindset (or evaluation culture) is potentially interesting and valuable across the organization.

- Executives and leadership in similar cultural institutions (museums, science centers, botanical gardens, operas, etc.).
- Staff at zoos and aquariums.
- Evaluators working with zoos and aquariums (and similar cultural organizations).
- Social science researchers working on similar topics.
- The Association of Zoos and Aquariums (AZA) and other professional support organizations.

For education directors at zoos and aquariums, work with evaluators is happening, but perhaps with mixed results in regards to its impact on department or organizational culture. To maximize these process benefits in zoo and aquarium settings and encourage the development of evaluative thinking and a strong evaluation culture; the study makes the following recommendations:

1. **Being intentional and expansive when planning work with evaluators can maximize benefits.** Most study respondents (90%) indicated work with professional evaluators in some capacity. Whether that is being mandated by grant funding or an opportunity presented by work with university partners or other peers, being intentional and expansive about who participates in these processes, and what they are expected to gain, may provide significant additional benefits to a workgroup's evaluation culture.

- Informing staff of upcoming evaluation efforts and how they can be involved increases transparency in communication and signals the value of participation.
- Considering a broader range of staff for participation in the planning processes (e.g., the development of logic models or evaluation questions) provides professional development opportunities. There may be some additional cost for an evaluator to broaden the scope of their work to include these educational opportunities, but these may be much less than training courses and provide practical and relevant examples.
- Providing opportunities for staff to serve as data collectors gives staff valuable practical training and experience with evaluation skills and may expand the scope of data collected.
- Posting evaluation results publicly and keeping a record of evaluations conducted and questions answered can help organizational learning be more constructive than cyclical. It may also help establish the growth mindset of a team among new staff. The lowest scoring item on the survey asked whether there were adequate records of past change efforts.

2. **Learning to recognize and reinforce examples of evaluative thinking in staff can help establish a team norm.** Recognizing and rewarding examples of social learning, reflective practice, interrogations of assumptions/positionality, and data-driven inquiry and decision-making among a team may help demonstrate that leadership and the organization value the development of these skills.

- Using a tool like the ROLE instrument (or the modified ROLE used in this study) to take the pulse of a team may allow a manager to both see progress and create an ongoing dialog with the team around these ideas.

3. **Being a vocal advocate and model for evaluative thinking in the workgroup and organization sets an example.** Leadership support was the primary antecedent positively associated with a strong evaluation culture (and a key influence on psychological safety). By providing resources (time, funding, expertise) and demonstrating evaluative thinking amongst the team, managers can lead through action.

    - Building time into a team's schedules for reflective practice and establishing it as an expectation gives permission and *time* that staff may feel is out of their authority. Leaders participating in reflection activities may further establish this practice as a team value.
    - Being clear about the assumptions and data behind managerial decisions and clearly communicating team expectations around using data and surfacing assumptions in staff decisions and planning may help to further establish evaluative thinking as common practice in a workgroup. This could be as simple as asking, "Why do you/we think that?" or "How do you know?"
    - Encouraging staff to seek each other out for problem-solving and to share how problems resolve in team settings (e.g., staff meetings) may help reinforce the social nature of evaluative thinking and prevent learning from being isolated.

4. **The psychological safety of a team should not be taken for granted.** If a team does not feel safe to ask questions, take risks, and make mistakes, then learning and growth are unlikely. Being explicit and intentional about creating a safe work environment lays the foundation for staff to challenge ideas, innovate and evolve.
    - Creating dialogue around the learning associated with mistakes and *failures* with staff individually and with the team (as appropriate) may work to reduce the stigma associated with mistakes.
    - For risk-averse staff, establishing a practice of talking through calculated risks with a supervisor or peer may help them understand the stakes and potential benefits or learnings.
    - When managers are conscious of responses (verbal and non-verbal) to ideas shared by staff that are different from their own, or organizational/ department norms, they demonstrate openness to ideas with actions as well as words.

Executives at zoos and aquariums (and other cultural institutions) hold the ultimate authority over budgets and the ultimate responsibility for organizational culture. Zoos and aquariums are investing in evaluation activities, but are they having the impact on their organizations that they could? Leadership was the most significantly associated antecedent of evaluation culture in the literature owing their ability to facilitate development through taking ownership for evaluation, establishing organizational values, investing in systems and structures that support learning, communicating clearly around decision-making, and providing resources, incentives and accountability (Coopey, 1995; Fleming & Easton, 2010; Jo

& Joo, 2011; Marsick & Watkins, 2003; Mayne, 2009; Murphy, 1999; Preskill & Boyle, 2008; Preskill & Torres, 1999b; Sanders, 2003; Stufflebeam, 2002; Taut, 2007; Taylor-Powell & Boyd, 2008; Volkov & King, 2007; Williams & Hawkes, 2003; and others). Executive level leaders can demonstrate both through their language and practices that the organization values evaluative thinking and a growth mindset. Investment in evaluation is one step, but executives may also consider investing in evaluative thinking by encouraging space in staff schedules for reflective practices, even if it means adjusted expectations about revenue or participation metrics. Creating more efficient or effective programs for slightly smaller audiences may eventually put those programs in a better position to scale up. Executives can also demonstrate a commitment to examining organizational assumptions and positionality by investing in diversity, equity, inclusion, and justice (DEIJ) programs.

The Association of Zoos and Aquariums requires evaluation of education programming, specifying outcome evaluation rather than just participation satisfaction surveys. Several interview respondents noted accreditation requirements as among the spurs to develop their evaluation practices. However, accreditation is required only every five years and accreditation inspection teams emphasize experience in animal care, safety, and overall organizational executive function. There are standards that inspectors can apply to evaluation-related accreditation materials, but the direct experience of inspectors in program evaluation is typically limited. Steps could be taken in the accreditation process to encourage broader adoption of evaluative thinking practices in addition to the conduct of evaluations. These could include adding more inspectors with direct experience with evaluation and/or organizational development; providing interview questions for inspectors to inquire on leadership support for evaluation and evaluative thinking, how decisions are made, and other elements of evaluative thinking; and updating accreditation standards to reflect evaluative thinking in addition to evaluations.

Interview respondents also noted peer pressure as a motivation for improving evaluation practices and culture. For zoos and aquariums, AZA is the medium for peer pressure across the industry. AZA already does an excellent job incorporating evaluation into conference programming, their publications, committee work, and in the association's professional development courses. Much of this work is focused on building evaluation skills, improving attitudes towards evaluation, and conducting evaluations. Including narratives and examples of how evaluation activities and working with evaluators has impacted the organizational culture of peer institutions could further enhance this movement, especially when it happens at the executive level.

Staff at zoos and aquariums may also be interested in the findings of this study, especially around the discussion of psychological safety and the recommendation to expand participation in evaluation activities. The conversation around psychological safety may give voice to concerns or limitations staff may be feeling in their work groups around openness to staff input, risk, and the consequences of making mistakes. Edmondson (1999) suggested that psychological safety went beyond trust to include feelings of mutual respect and care for team members. It may be that what staff feel in their work group is less about trust as that desire to be respected by supervisors and peers for the value they bring to their collective work. Having this language may lead to more productive conversations with peers and managers.

Staff should also be proactive in expressing interest to their managers and supervisors about participation in evaluation activities. Managers may be reluctant to include staff believing they are already feeling busy with existing work. While it does require supervisors making space for this participation, knowing that staff are interested can provide motivation for providing this opportunity. Staff should also hold management accountable for posting and archiving evaluation results so that the learning is socialized amongst the team rather than isolated in a few staff members' experience. Managers may again assume that staff are too busy or uninterested in reviewing evaluation results. Creating a shared understanding amongst staff and management of the value and interest in participation can help each group challenge their own assumptions, develop shared accountability, and work together on solutions.

While the survey results for this study did not see a relationship between work with professional evaluators and a workgroups evaluation culture, the information regarding *how* institutions work with evaluators was limited to the nine case-study interviews. Survey questions included program evaluators with audience researchers and university partnerships although each category of evaluator might reasonably have a different kind of influence on evaluative thinking or culture. Only 15% of respondents indicated internal program evaluator staff, though 70% indicated project work with program evaluators. Though they may have the potential for the deepest and most extensive engagement with leadership and staff, program evaluators were also noted by interview respondents as sometimes/ often limited to interactions with organizational leadership and/or select program staff. Existing scholarship routinely describes the potential of program evaluators to serve as educators and coaches (Beere, 2005; Cousins et al., 2014; Garcia-Iriarte et al., 2011; Volkov, 2011). Work with program evaluators could represent the greatest opportunity to influence the evaluation cultures of zoos and aquariums if internal program evaluators are empowered by leaders towards culture development and project work with external program evaluators can be expanded to include greater staff participation and professional development.

Recommendations for Future Research

This study represents a first step at looking at the relationship between professional evaluators and organizational evaluation culture and presents several potentially interesting extensions and new directions.

One survey reviewer emphasized that it was important to understand not only that organizations were working with evaluators, but *how* they were working with evaluators. This was borne out in the case-study interviews. A binary presence/absence condition for internal evaluation staff could include both the established evaluation and research department of Institution E and the brand-new evaluation coordinator/facilitator of Institution C. Pertinent research questions might include: is evaluation capacity-building part of the job descriptions of internal program evaluators; do the agreed upon outcomes of project-based external evaluation contracts include staff/culture development activities, how extensively are staff involved in evaluation activities, and how are results of evaluations reported and archived? If a strong relationship between work with evaluators and evaluation culture had been established in the survey results, the next logical questions would have been around why, and how. With no relationship evident, despite support in interviews and the literature for a positive association, the question becomes, why not?

The ROLE instrument was a good fit for the study constructs established for this research. However, an alternative approach may have created an instrument that was a better fit, in retrospect. By starting with a set of focus groups or interviews with key-informants (education directors, researchers, evaluators), the study would have proceeded with fewer assumptions about the extent of factors associated with or supportive of an evaluation culture. Professional events like the AZA midyear and annual conferences provided opportunities for roundtable engagements that could have also included a diverse group of managers, staff, and evaluators. Starting with this qualitative approach may have surfaced constructs like psychological safety or suggested factors, like the background of education directors, that would have been relevant to include in a survey instrument. A longer series of follow-up interviews may also uncover more nuance regarding the details of how professional evaluators work in zoo and aquarium settings. A study designed in this way may still be useful to explore other factors influencing the development of evaluation culture.

One potentially interesting factor to explore would be the influence of the education director's background on their team's evaluation culture. The influence of leadership actions has been well established, but what motivates or inspires those actions? All education directors in the interview sample indicated limited formal training in evaluation, with most of their education and experience coming from on-the-job experience and conference- or association-related workshops. The education director from Institution D previously worked at several institutions that had either invested in internal evaluation capacity or for supervisors that demonstrated value for evaluation through language or actions. Institution D scored at the mean of the sample, but was an example of an institution that had created a system where external evaluators were consistently active in the activities of the department through their involvement in capital projects.

When reviewing ideas linked to the idea of fostering an evaluation culture from the academic literature, 18 different concepts were mentioned more than twice (see Table 2). Only three of these (transparent communication, openness to change, risk-taking) showed any overlap with constructs related to psychological safety, but many seem to take the psychological safety of staff for granted (staff ask questions, communities of practice). While there is ample research around the role of psychological safety in the workplace, there seems to be none that specifically links psychological safety to the development of evaluative thinking or an evaluation culture within a workgroup or organization. Judging from the responses by case study interviewees, there is likely significant interest in participating in such work.

Interview subjects were mixed on the question of whether staff would respond to the survey questions similarly. Three of the nine thought they would generate scores comparable to their own, but the remaining six felt scores would be different with differences highlighted by staff's tenure with the organization and their level of involvement with evaluation efforts . . . though participants did not agree whether more involvement would make staff more supportive or more critical of current efforts. While leadership support is a critical aspect of both the development of an evaluation culture and in establishing psychological safety in the workgroup, understanding how they are aligned with staff views is important. Leadership may be limited by their connection to day-to-day activities and possess biases or limitations in how objectively they can view leadership actions and contributions to culture. For a full and complete view of the evaluation culture in a team or organization, the perspectives of both leadership and staff are necessary.

This study was delimited to the education or programming departments at zoos and aquariums, but evaluative thinking and a culture of evaluation returns benefits across any organization. This is well supported by the extensive literature on the development of learning organizations. In this study, a third of respondents that indicated the presence of internal evaluation staff also indicated that internal staff worked in a different department (or there were additional staff that worked outside the department). Reflective practices, social learning, challenging assumptions, and data-driven decision-making are relevant well beyond education departments. A modified study design that included leadership (and/or staff) from across organizations could reveal interesting relationships between the components of evaluative thinking and an organization's evaluation culture.

Finally, since institutional size (as measured by operating budget and annual attendance) was positively associated with investment in evaluation, it would be interesting to look more closely at the variation in responses within institutional size categories by recruiting more institutions within each size category, and possibly expanding to include the ultra-large organizations. Alternatively, additional institutions could be recruited from within the existing sample for follow-up interviews.

References

Arnold, M. E. (2006). Developing evaluation capacity in extension 4-H Field faculty: A framework for success. *American Journal of Evaluation*, *27*(2), 257–269.

Association of Zoos & Aquariums. (n.d.). About us. Retrieved from https://www.aza.org/about-us

Association of Zoos & Aquariums. (2018a). Benchmark reports. Retrieved from https://www.aza.org/benchmark-reports

Association of Zoos & Aquariums. (2018b). Zoo and aquarium statistics. Retrieved from https://www.aza.org/zoo-and-aquarium-statistics

Association of Zoos & Aquariums. (2019). *Accreditation standards & related policies, 2019 edition*. Silver Spring, MD: Author.

Baker, A., & Bruner, B. (2006). *Evaluation capacity and evaluative thinking in organizations.* Cambridge, MA: Bruner Foundation.

Barnette, J. J., Wallis, A. B., & Barber Wallis, A. (2003). Helping evaluators swim with the current: training evaluators to support mainstreaming. *New Directions for Evaluation*, *2003*(99), 51–61.

Beere, D. (2005). Evaluation capacity-building: A tale of value-adding. *Evaluation Journal of Australasia*, *5*(2), 41-47.

Botcheva, L., White, C. R., & Huffman, L. C. (2002). Learning culture and outcomes measurement practices in community agencies. *American Journal of Evaluation*, *23*(4), 421–434.

Brewer, C. (2001). Cultivating conservation literacy: “Trickle-down” education is not enough. *Conservation Biology*, *15*(5), 1203-1205.

Brown, R. (1990). Politeness theory: Exemplar and exemplary. In I. Rock (Ed.), *The legacy of Solomon Asch: Essays in cognition and social psychology* (pp. 23–38). New York, NY: Psychology Press.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, *36*(3), 375–388.

Carman, J. G., & Fredericks, K. A. (2010). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation*, *31*(1), 84–104.

Carmeli, A., Brueller, D., & Dutton, J. E. (2009). Learning behaviours in the workplace: The role of high-quality interpersonal relationships and psychological safety. *Systems Research and Behavioral Science*, *26*(1), 81–98.

Clavijo, K., Fleming, M. L., Hoermann, E. F., Toal, S. A., & Johnson, K. (2005). Evaluation use in nonformal education settings. *New Directions for Evaluation*, *2005*(108), 47–55.

Coopey, J. (1995). The learning organization, power, politics and ideology introduction. *Management Learning*, *26*(2), 193–213.

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, *2014*(141), 7–23.

De Peuter, B., & Pattyn, V. (2009). Evaluation capacity: Enabler or exponent of evaluation culture? In *Policy and Programme Evaluation in Europe: Cultures and Prospects* (pp. 133–142). Paris, France: l'Harmattan.

Duignan, P. (2003). Mainstreaming evaluation or building evaluation capability? Three key elements. *New Directions for Evaluation*, *2003*(99), 7-21.

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 23–43.

Edmondson, A. C., & Moingeon, B. (1998). From organizational learning to the learning organization. *Management Learning*, *29*(1), 5–20.

Ewell, P. T. (2002). A delicate balance: The role of evaluation in management. *Quality in Higher Education*, *8*(2), 159–171.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, *26*(2), 132–139.

Fierro, L. A., Codd, H., Gill, S., Pham, P. K., Targos, P. T. G., & Wilce, M. (2018). Evaluative thinking in practice: The National Asthma Control Program. *New Directions for Evaluation*, *2018*(158), 49–72.

Fleming, M. L., & Easton, J. (2010). Building environmental educators' evaluation capacity through distance education. *Evaluation and Program Planning*, *33*(2), 172–177.

Fraser, J., & Sickler, J. (2009). Measuring the cultural impact of zoos and aquariums. *International Zoo Yearbook*, *43*(1), 103–112.

Fraser, J., & Wharton, D. (2007). The future of zoos: A new model for cultural institutions. *Curator: The Museum Journal*, *50*(1), 41–54.

Fraser, J., Heimlich, J. E., Ogden, J., Atkins, A., McReynolds, S., Chen, C., . . . Boyle, *P*. (2010). *The AZA's framework for zoo and aquarium social science research.* Silver Spring, MD: Association of Zoos & Aquariums.

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological safety: A meta-analytic review and extension. *Personnel Psychology*, *70*(1), 113–165.

García-Iriarte, E., Suarez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2011). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation*, *32*(2), 168–182.

Gibbs, G. R. (2007). Thematic coding and categorizing. *Analyzing Qualitative Data*, *703*, 38–56.

Grudens-Schuck, N. (2003). The rigidity and comfort of habits: A cultural and philosophical analysis of the ups and downs of mainstreaming evaluation. *New Directions for Evaluation*, *2003*(99), 23-32.

Hargreaves, M. B., & Podems, D. (2012). Advancing systems thinking in evaluation: A review of four publications. *American Journal of Evaluation*, *33*(3), 462–470.

Heimlich, J. E., & Horr, E. E. T. (2010). Adult learning in free-choice, environmental settings: What makes it different? *New Directions for Adult and Continuing Education*, *2010*(127), 57–66.

Hobson, D. (2001). Action and reflection: Narrative and journaling in teacher research. In Burnaford, G., Fischer, J., & Hobson, D. (Eds.), *Teachers Doing Research* (pp. 7-27). Mahwah, NJ: Lawrence Erlbaum Associates.

Hueftle Stockdill, S., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, *2002*(93), 7–26.

Jenks, B., Vaughan, P. W., & Butler, P. J. (2010). The evolution of Rare Pride: Using evaluation to drive adaptive management in a biodiversity conservation organization. *Evaluation and Program Planning*, *33*(2), 186–190.

Jo, S. J., & Joo, B. K. (2011). Knowledge sharing: The influences of learning organization culture, organizational commitment, and organizational citizenship behaviors. *Journal of Leadership & Organizational Studies*, *18*(3), 353–364.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, *33*(4), 692–724.

Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The effect of invitation design on web survey response rates. *Social Science Computer Review*, *30*(3), 339–349.

Khalil, K., & Ardoin, N. (2011). Programmatic evaluation in Association of Zoos and Aquariums-accredited zoos and aquariums: A literature review. *Applied Environmental Education & Communication*, *10*(3), 168–177.

Kubarek, J. (2015). Building staff capacity to evaluate in museum education. *Journal of Museum Education*, *40*(1), 8–12.

Kubarek, J., & Trainer, L. (2015). Empowering museum educators to evaluate. *Journal of Museum Education*, *40*(1), 3–7.

Kular, S., Gatenby, M., Rees, C., Soane, E., & Truss, K. (2008). *Employee engagement: A literature review*. London, United Kingdom: Kingston Business School, Kingston University.

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, *33*(3), 307–338.

Luebke, J. F., & Grajal, A. (2011). Assessing mission-related learning outcomes at zoos and aquaria: Prevalence, barriers, and needs. *Visitor Studies*, *14*(2), 195–208.

Marsick, V. J., & Watkins, K. E. (2003). Demonstrating the value of an organization's learning culture: The dimensions of the learning organization questionnaire. *Advances in Developing Human Resources*, *5*(2), 132–151.

Matiasek, J., & Luebke, J. F. (2014). Mission, messages, and measures: Engaging zoo educators in environmental education program evaluation. *Studies in Educational Evaluation*, *41*, 77–84.

Mayne, J. (2008). *Building an evaluative culture for effective evaluation and results management*. Rome, Italy: Insitutional Learning and Change (ILAC) Initative.

Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, *24*(2), 1–30.

Monroe, M. C., & Adams, D. C. (2012). Increasing response rates to web-based surveys. *Journal of Extensions*, *50*(6), 6-7.

Monroe, M. C., Fleming, M. L., Bowman, R. A., Zimmer, J. F., Marcinkowski, T., Washburn, J., & Mitchell, N. J. (2005). Evaluators as educators: Articulating program theory and building evaluation capacity. *New Directions for Evaluation*, *2005*(108), 57–71.

Muñoz-Leiva, F., Sánchez-Fernández, J., Montoro-Ríos, F., & Ibáñez-Zapata, J. Á. (2010). Improving the response rate and quality in Web-based surveys through the personalization and frequency of reminder mailings. *Quality and Quantity*, *44*(5), 1037–1052.

Murphy, D. F. (1999). Developing a culture of evaluation. *The Journal of TESOL France*, *6*, 5-13.

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review*, *27*(3), 521–535.

Ogden, J., & Heimlich, J. E. (2009). Why focus on zoo and aquarium education? *Zoo Biology*, *28*(5), 357–360.

Ortenbiad, A. (2002). A typology of the idea of learning organization. *Management Learning*, *33*(2), 213–230.

Ose, S. O. (2016). Using Excel and Word to structure qualitative data. *Journal of Applied Social Science*, *10*(2), 147–162.

Owen, J. M. (2003). Evaluation culture: A definition and analysis of its development within organisations. *Evaluation Journal of Australasia*, *3*(1), 43–47.

Owen, J. M., & Lambert, F. C. (1995). Roles for evaluation in learning organizations. *Evaluation*, *1*(2), 237–250.

Owen, K., & Visscher, N. (2015). Museum-university collaborations to enhance evaluation capacity. *Journal of Museum Education*, *40*(1), 70–77.

Patton, M. Q. (1998). Discovering process use. *Evaluation*, *4*(2), 225–233.

Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.

Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. *New Directions for Evaluation*, *2018*(158), 11–28.

Picciotto, R. (2013). Evaluation independence in organizations. *Journal of Multidisciplinary, 9*(20), 18–32.

Porter, S. R., & Whitcomb, M. E. (2016). The impact of contact type on web survey response rates. *The Public Opinion Quarterly*, *67*(4), 579–588.

Preskill, H. (2008). Evaluation's second act: A spotlight on learning. *American Journal of Evaluation*, *29*(2), 127–138.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, *29*(4), 443–459.

Preskill, H., & Torres, R. T. (1999a). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.

Preskill, H., & Torres, R. T. (1999b). Building capacity for organizational learning through evaluative inquiry. *Evaluation*, *5*(1), 42–60.

Preskill, H., & Torres, R. T. (2000a). The learning dimension of evaluation use. *New Directions for Evaluation*, *2000*(88), 25–37.

Preskill, H., & Torres, R. T. (2000b). The readiness for organizational learning and evaluation instrument (ROLE). *Evaluation in Organizations*, 421–434.

Preskill, H., & Zuckerman, B. (2003). An exploratory study of process use. *American Journal of Evaluation*, *24*(4), 423–442.

Preskill, H., Torres, R., & Martinez-Papponi, B. (1999, November). Assessing an organization's readiness for learning from evaluative inquiry. In *American Evaluation Association Annual Conference*. Orlando, FL.

Rabb, G. B. (2004). The evolution of zoos from menageries to centers of conservation and caring. *Curator: The Museum Journal*, *47*(3), 237–246.

Roe, K., Mcconney, A., & Mansfield, C. (2014). Using evaluation to prove or to improve ? An international, mixed method investigation into zoos' education evaluation practices. *Journal of Zoo and Aquarium Research, 2*(4), 108–116.

Sanders, J. R. (2002). Presidential address: On mainstreaming evaluation. *American Journal of Evaluation*, *23*(3), 253–259.

Sanders, J. R. (2003). Mainstreaming evaluation. *New Directions for Evaluation*, *2003*(99), 3–6.

Sanner, B., & Bunderson, J. S. (2015). When feeling safe isn't enough: Contextualizing models of safety and learning in teams. *Organizational Psychology Review*, *5*(3), 224–243.

Sauermann, H., & Roach, M. (2013). Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Research Policy*, *42*(1), 273–286.

Schein, E. H. (1993). How can organizations learn faster? The challenge of entering the green room. *Sloan Management Review*, *34*, 85–92.

Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. *New Directions for Evaluation*, *2018*(158), 125–137.

Scriven, M. S. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.

Seattle Aquarium Society. (2011). *Strategic Plan 2011-2030*. Seattle, WA: Author.

Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of Computer-Mediated Communication*, *6*(2), JCMC621.

Somers, C. (2005). Evaluation of the Wonders in Nature–Wonders in Neighborhoods conservation education program: Stakeholders gone wild! *New Directions for Evaluation*, *2005*(108), 29–46.

Steele-Inama, M. (2015). Building evaluation capacity as a network of museum professionals. *Journal of Museum Education*, *40*(1), 78–85.

Stufflebeam, D. L. (2002). Institutionalizing evaluation checklist [PDF file]. Retrieved from http://www. wumich.edu/evalctr/checklists

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass.

Suárez-Herrera, J. C., Springett, J., & Kagan, C. (2009). Critical connections between participatory evaluation, organizational learning and intentional change in pluralistic organizations. *Evaluation*, *15*(3), 321–342.

Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation*, *28*(1), 45–59.

Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation*, *2008*(120), 55–69.

Trouteaud, A. R. (2004). How you ask counts: A test of Internet-related components of response rates to a web-based survey. *Social Science Computer Review*, *22*(3), 385–392.

Uyen Tran, L., & King, H. (2007). The professionalization of museum educators: The case in science museums. *Museum Management and Curatorship*, *22*(2), 131–149.

Vo, A. T., & Archibald, T. (2018). New directions for evaluative thinking. *New Directions for Evaluation*, *2018*(158), 139–147.

Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of evaluative thinking. *New Directions for Evaluation*, *2018*(158), 29–47.

Volkov, B. B. (2011). Beyond being an evaluator: The multiplicity of roles of the internal evaluator. *New Directions for Evaluation*, *2011*(132), 25–42.

Volkov, B. B., & King, J. A. (2007). A checklist for building organizational evaluation capacity [PDF file]. Retrieved from http://www. wumich.edu/evalctr/checklists

Wandersman, A. (2014). Moving forward with the science and practice of evaluation capacity building (ECB): The why, how, what, and outcomes of ECB. *American Journal of Evaluation*, *35*(1), 87–89.

Wehipeihana, N., & McKegg, K. (2018). Values and culture in evaluative thinking: Insights from Aotearoa, New Zealand. *New Directions for Evaluation*, *2018*(158), 93–107.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation ? *American Journal of Evaluation*, *19*(1), 21–33.

Williams, D. D., & Hawkes, M. L. (2003). Issues and practices related to mainstreaming evaluation: Where do we flow from here? *New Directions for Evaluation*, *2003*(99), 63–83.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (2004). *Program evaluation: Alternative approaches and practical guidelines*. Boston, MA: Allyn and Bacon.

Zajkowska, M. (2012). Employee engagement: How to improve it through internal communication. *Human Resources Management & Ergonomics*, *6*(1), 104-117.

Appendix A: Informed Consent Language

Consent language included in the original survey was approved by USF's Institutional Review Board and covers both phases of the project. Formatting preserved.

Overview: You are being asked to take part in a research study. The following information should help you to decide if you would like to participate. The sections in this Overview provide basic information about the study. More detailed information follows.

Study Staff: This study is being led by Jim Wharton, a doctoral student at the University of South Florida. This person is called the Principal Investigator (PI). He is being guided in this research by Drs. Liliana Rodriguez-Campos and Robert Dedrick of the University of South Florida.

Study Details: The purpose of the study is to understand how working with professional evaluators might be related to the strength of a zoo or aquarium's evaluation culture. To address this question the PI will survey the education/engagement directors at AZA-accredited facilities. A subset of respondents will be invited to participate in a follow-up interview in the second phase of the study.

Participants: You are being asked to take part because you are the senior manager of the education/engagement department of your institution.

Voluntary Participation: Your participation is voluntary. You do not have to participate and may stop your participation at any time. Participating in phase one of this study **does not** obligate you to participate in phase two. There will be no penalties if you do not participate or decide to stop once you start.

Benefits, Compensation, and Risk: You will not be compensated monetarily for your participation. However, you may consider seeing your results in comparison to the sample to be valuable. These will be provided to you at a time after the survey analysis is complete. This research is considered minimal risk. Minimal risk means that study risks are the same as the risks you face in daily life.

Confidentiality: Even if we publish the findings from this study, we will keep your study information private and confidential.

Why are you being asked to take part?

This study includes U.S.-based AZA-accredited zoos and aquariums. Accredited institutions are required to evaluate their programming. Focusing on U.S.-based facilities will minimize language and cultural differences. The study is further delimiting participation to the education/engagement departments as this is where the evaluation need and/or function commonly resides.

Study Procedures

If you take part in this study, you will be asked to complete an online survey with 50 Likert-style questions and a short set of multiple choice and open-ended questions related to institutional demographics and your department's work with professional evaluators. The survey should take 15-20 minutes to complete. A subset of participants who complete the survey will be invited to participate in phase two of the study.

Phase two consists of a review of institutional documents related to evaluation, and an interview with the PI. These documents are being requested to provide context and a deeper understanding of the program and institution in advance of the interview. The documents will be read by the PI to help craft the interview questions and provide more detailed background in advance of the interview. These documents may be restricted to recent accreditation materials or may include additional documents of the participant's choosing. Interviews will be conducted via video conference platform (Zoom or similar) with questions provided in advance. Participants in phase two will receive a short report with an analysis of their survey results and recommendations based on the results of the interview (described in the 'Benefits and Risks' section below).

Alternatives / Voluntary Participation / Withdrawal

You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part. You are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study. Study results and conclusions will be made available via AZA conference poster or presentation and submitted for publication. You can also contact the PI directly for a copy of the completed dissertation.

Benefits and Risks

We are unsure if you will receive any benefits by taking part in this research study. This research is considered to be minimal risk. Seeing your results in comparison to the study sample may be beneficial to your work or professional development.

Compensation

You will be not be compensated for participating in this study.

Privacy and Confidentiality

We will do our best to keep your records private and confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Certain people may need to see your study records. The only people who will be allowed to see these records are: the Principle Investigator, listed faculty advisors, and The University of South Florida Institutional Review Board (IRB).

Your information collected as part of the research, even if identifiers are removed, will NOT be used or distributed for future research studies. It is possible, although unlikely, that unauthorized individuals could gain access to your responses because you are responding online. Confidentiality will be maintained to the degree permitted by the technology used. No guarantees can be made regarding the interception of data sent via the Internet. However, your participation in this online survey involves risks similar to a person's everyday use of the Internet.

Contact Information

If you have any questions, concerns, or complaints about this study, contact Jim Wharton at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

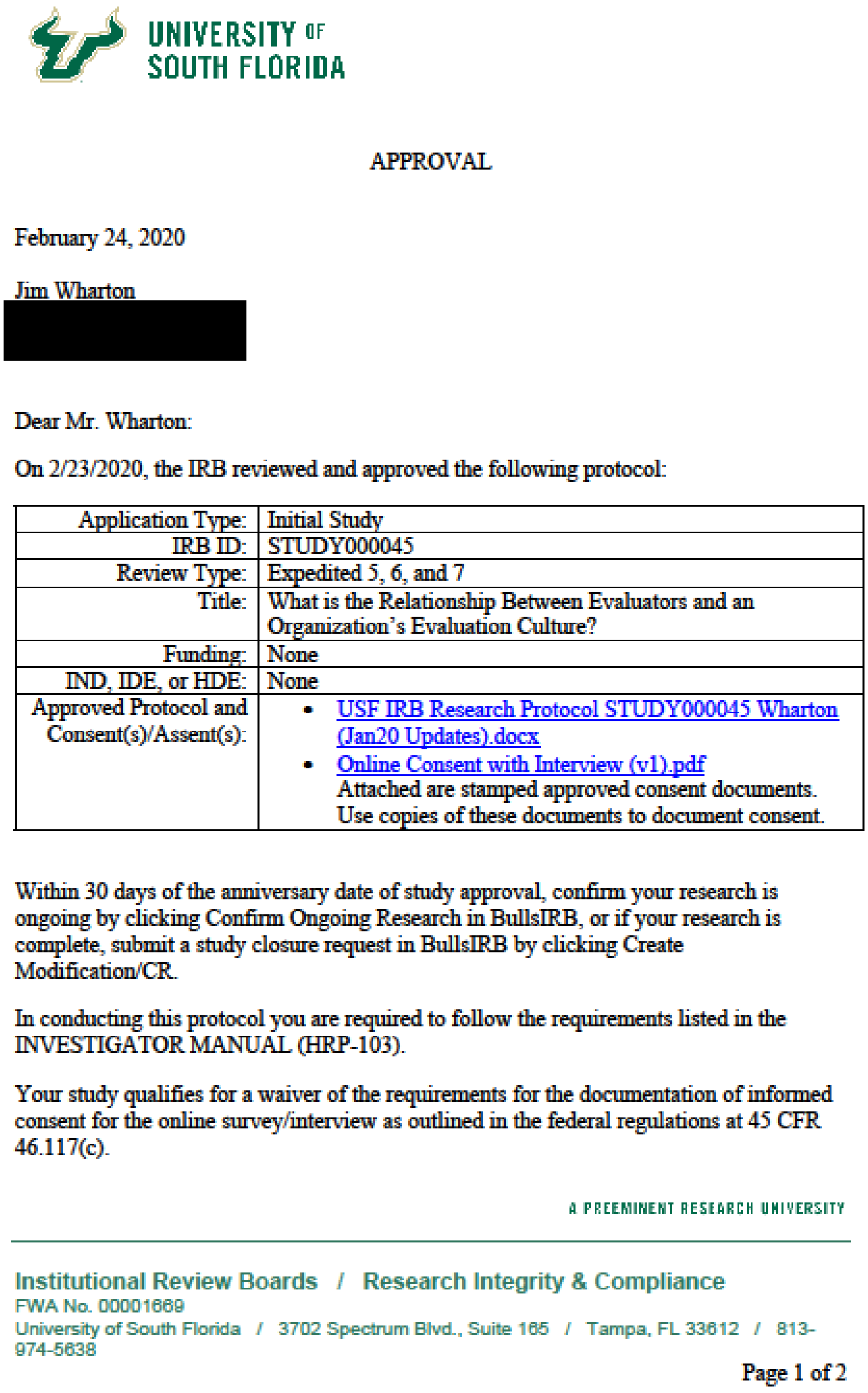
We may publish what we learn from this study. If we do, we will not let anyone know your name. We will not publish anything else that would let people know who you are. You can print a copy of this consent form for your records.

☐ I work for a zoo, aquarium, or other live-animal facility currently accredited by the Association of Zoos and Aquariums.

☐ I understand that by proceeding with this survey that I am agreeing to take part in research, and I am 18 years of age or older.

☐ I agree that, if I elected in the survey to be a candidate for phase two, the PI can contact me using the contact information I provided in the survey.

Appendix B: Institutional Review Board Approval



UNIVERSITY OF SOUTH FLORIDA

APPROVAL

February 24, 2020

Jim Wharton
[redacted]

Dear Mr. Wharton:

On 2/23/2020, the IRB reviewed and approved the following protocol:

Application Type:	Initial Study
IRB ID:	STUDY000045
Review Type:	Expedited 5, 6, and 7
Title:	What is the Relationship Between Evaluators and an Organization's Evaluation Culture?
Funding:	None
IND, IDE, or HDE:	None
Approved Protocol and Consent(s)/Assent(s):	• USF IRB Research Protocol STUDY000045 Wharton (Jan20 Updates).docx • Online Consent with Interview (v1).pdf Attached are stamped approved consent documents. Use copies of these documents to document consent.

Within 30 days of the anniversary date of study approval, confirm your research is ongoing by clicking Confirm Ongoing Research in BullsIRB, or if your research is complete, submit a study closure request in BullsIRB by clicking Create Modification/CR.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Your study qualifies for a waiver of the requirements for the documentation of informed consent for the online survey/interview as outlined in the federal regulations at 45 CFR 46.117(c).

A PREEMINENT RESEARCH UNIVERSITY

Institutional Review Boards / Research Integrity & Compliance
FWA No. 00001669
University of South Florida / 3702 Spectrum Blvd., Suite 165 / Tampa, FL 33612 / 813-974-5638

Page 1 of 2

UNIVERSITY OF
SOUTH FLORIDA

Sincerely,

Various Menzel
IRB Research Compliance Administrator

A PREEMINENT RESEARCH UNIVERSITY

Institutional Review Boards / Research Integrity & Compliance
FWA No. 00001669
University of South Florida / 3702 Spectrum Blvd., Suite 165 / Tampa, FL 33612 / 813-974-5638

Page 2 of 2

Appendix C: Original ROLE Survey Instrument

In this section you will find the original version of the Readiness for Organizational Learning and Evaluation (ROLE) survey instrument (Preskill & Torres, 2000b). All questions had a 5-point Likert-style response scale ranging from strongly disagree at 1 to strongly agree at 5. Questions 60-62 are on a binary, yes/no scale.

Table C1

Original ROLE Survey Instrument (Preskill & Torres, 2000b)

Question	**Organizational Culture**[a]
	Culture and Problem Solving[b]
1	Employees respect each other's perspectives and opinions.
2	Employees ask each other for information about work issues and activities.
3	Employees continuously look for ways to improve processes, products and services.
4	Employees are provided opportunities to think about and reflect on their work.
5	Employees often stop to talk about the pressing work issues we're facing.
6	When trying to solve problems, employees use a process of working through the problem before identifying solutions.
7	There is little competition among employees for recognition or rewards.
8	Employees operate from a spirit of cooperation, rather than competition.
9	Employees tend to work collaboratively with each other.
10	Employees are more concerned about how their work contributes to the success of the organization than they are about their individual success.
11	Employees face conflict over work issues in productive ways.
12	Employees generally view problems or issues as opportunities to learn.
	Risk Taking
13	Mistakes made by employees are viewed as opportunities for learning.
14	Employees continuously ask themselves how they're doing, what they can do better, and what is working.
15	Employees are willing to take risks in the course of their work.
16	Employees are committed to being innovative and forward looking.
17	Employees are confident that mistakes or failures will not affect them negatively.
	Participatory Decision Making
18	Employees generally trust their managers or supervisors.

Table C1 (continued)

Question	**Organizational Culture (continued)**
19	Managers and supervisors view individuals' capacity to learn as the organization's greatest resource.
20	Employees use data/information to inform their decision-making.
21	Asking questions and raising issues about work is encouraged.
22	Employees are not afraid to share their opinions even if those opinions are different from the majority.
23	I feel safe explaining to others why I think or feel the way I do about an issue.
24	Employees are encouraged to take the lead in initiating change or in trying to do something different.
25	Managers and supervisors make decisions after considering the input of those affected.
26	In meetings employees are encouraged to discuss the values and beliefs that underlie their opinions.
27	Employees are encouraged to offer dissenting opinions and alternative viewpoints.
	Leadership
28	Managers and supervisors admit when they don't know the answer to a question.
29	Managers and supervisors take on the role of coaching, mentoring and facilitating employees' learning.
30	Managers and supervisors help employees understand the value of experimentation and the learning that can result from such endeavors.
31	Managers and supervisors make realistic commitments for employees (e.g., time, resources, workload).
32	Managers and supervisors understand that employees have different learning styles and learning needs.
33	Managers and supervisors are more concerned with serving the organization than with seeking personal power or gain.
34	Managers and supervisors are open to negative feedback from employees.
35	Managers and supervisors model the importance of learning through their own efforts to learn.
36	Managers and supervisors believe that our success depends upon learning from daily practices.
37	Managers and supervisors support the sharing of knowledge and skills among employees.
38	Managers and supervisors provide the necessary time and support for systemic, long-term change.
39	Managers and supervisors use data/information to inform their decision-making.
	Systems and Structures
	Open and Accessible Work Environment
40	There is little bureaucratic red tape when trying to do something new or different.
41	Workspaces are designed to allow for easy and frequent communication with each other.
42	There are few boundaries between departments/units that keep employees from working together.
43	Employees are available (i.e., not out of the office or otherwise too busy) to participate in meetings.

Table C1 (continued)

Question	Systems and Structures (continued)
	Rewards and Recognition Systems
44	Employees are recognized or rewarded for learning new knowledge and skills.
45	Employees are recognized or rewarded for helping solve business/organizational problems.
46	The current reward or appraisal system recognizes, in some way, team learning and performance.
47	Employees are recognized or rewarded for helping each other learn.
48	Employees are recognized or rewarded for experimenting with new ideas.
	Relationship of Work to Organizational Goals
49	Employees understand how their work relates to the goals or mission of the organization.
50	Employees' performance goals are clearly aligned with the organization's strategic goals.
51	Employees meet work deadlines.
	Communications
	Availability
52	Information is gathered from clients, customers, suppliers or other stakeholders to gauge how well we're doing.
53	Currently available information tells us what we need to know about the effectiveness of our programs, processes, products, and services.
54	There are adequate records of past change efforts and what happened as a result.
	Dissemination
55	There are existing systems to manage and disseminate information for those who need and can use it.
56	Employees are cross trained to perform various job functions.
57	Employees have access to the information they need to make decisions regarding their work.
58	Employees use technologies to communicate with one another.
59	When new information that would be helpful to others is learned or discovered, it gets disseminated to those individuals.
	Teams
60	My department/unit currently operates via (or is transitioning towards) a team-based structure.
61	Employees are provided training on how to work as a team member.
62	My work is sometimes conducted as part of a working group that is or could be identified as a "team."
63	When conflict arises among team members, it is resolved effectively.
64	Team members are open and honest with one another.
65	Team meetings are well facilitated.
66	Team meetings address both team processes and work content.
67	Team meetings strive to include everyone's opinion.
68	Teams are encouraged to learn from each other and to share their learning with others.
69	Teams accomplish work they are charged to do.
70	Teams are an effective way to meet an organization's goals.

Table C1 (continued)

Question	**Evaluation**
71	The integration of evaluation activities into our work has enhanced (or would enhance) the quality of decision-making.
72	It has been (or would be) worthwhile to integrate evaluation activities into our daily work practices.
73	Managers and supervisors like (or would like) us to evaluate our efforts.
74	Evaluation helps (or would help) us provide better programs, processes, products and services.
75	There would be support among employees if we tried to do more (or any) evaluation work.
76	Doing (more) evaluation would make it easier to convince managers of needed changes.
77	This would be a good time to begin (or renew or intensify) efforts to conduct evaluations.
78	There are evaluation processes in place that enable employees to review how well changes we make are working.

Note. [a]These headings represent the original six dimensions of the instrument. [b]These secondary headings (in italics) represent subscales present in some dimensions.

Appendix D: Modified ROLE Survey Instrument

This section contains the modified ROLE instrument, as distributed. All 50 items are on a 0-100 scale (strongly disagree to strongly agree). Questions related to institutional demographics and work with professional evaluators follow.

Introductory Text

The purpose of this study is to understand how working with professional evaluators might be related to the strength of a zoo or aquarium's evaluation culture. This instrument will help you characterize the evaluation culture of your department. In some cases, you will be asked to judge the opinions or feelings of your department employees to the best of your ability.

You will be asked to respond to 50 items on a visual analog scale (similar to a Likert scale) where 0 indicates strong disagreement and 100 indicates total agreement. There will also be 14 multiple choice and open-ended questions about your institution and its work with professional evaluators. The survey will take you 15 to 20 minutes to complete.

For the statements that follow, this survey uses bars (instead of more typical Likert choices) so that respondents can be more precise. Click or touch the place along the bar that indicates your level of agreement. Your 'score' (0-100) will show on the left. You may tweak or change your score as often as you like.

Table D1

Modified ROLE Survey Instrument Used in Study

Question	**Organizational Culture**[a]
1	Department employees respect each other’s perspectives and opinions.
2	Department employees ask each other for information about work issues and/or activities.
3	Department employees continuously look for ways to improve processes, products and/or services.
4	Department employees are provided opportunities to think about and reflect on their work.
5	Department employees often stop to talk with each other about the pressing work issues we’re facing.
6	When trying to solve problems, department employees use a process of working through the problem before identifying solutions.
7	Department employees operate from a spirit of cooperation, rather than competition.
8	Department employees tend to work collaboratively with each other.
9	Mistakes made by department employees are viewed as opportunities for learning.
10	Department employees continuously ask themselves how they’re doing, what they can do better, and what is working.
11	Department employees are confident that mistakes or failures will not affect them negatively.
12	Managers and supervisors in the department view individuals’ capacity to learn as among the organization’s greatest resources.
13	Department employees use data/information to inform their decision-making.
14	Asking questions and raising issues about work with department leaders is encouraged.
15	Department employees are not afraid to share their opinions in meetings, even if those opinions are different from the majority.
16	Department employees feel safe explaining to others why they think or feel the way they do about an issue.
17	Department employees are encouraged to take the lead in initiating change or in trying to do something different.
18	Managers and supervisors in the department make decisions after considering the input of those affected.
19	In meetings, department employees are encouraged to discuss the values and beliefs that underlie their opinions.
20	Department employees are encouraged to offer dissenting opinions and alternative viewpoints.
	Leadership
21	Managers and supervisors in the department take on the role of coaching, mentoring and facilitating employees’ learning.
22	Managers and supervisors in the department help employees understand the value of experimentation and the learning that can result from such endeavors.
23	Managers and supervisors in the department are open to negative feedback from employees.

Table D1 (continued)

Question	**Leadership (continued)**
24	Managers and supervisors in the department model the importance of learning through their own efforts to learn.
25	Managers and supervisors in the department believe that success depends upon learning from daily practices.
26	Managers and supervisors in the department support the sharing of knowledge and skills among employees.
27	Managers and supervisors in the department provide the necessary time and support for systemic, long-term change.
28	Managers and supervisors in the department use data/information to inform their decision-making.
	Systems & Structures
29	There is little bureaucratic red tape when trying to do something new or different in the department.
30	There are few boundaries between department units or working groups that keep employees from working together.
31	Department employees are recognized or rewarded for learning new knowledge and skills.
32	Department employees are recognized or rewarded for helping solve organizational problems.
33	The current reward or appraisal system in the department recognizes, in some way, team learning and performance.
34	Asking questions and raising issues about work with department leaders is encouraged.
35	Department employees are recognized or rewarded for experimenting with new ideas.
	Communication of Information
36	Information is gathered from guests, program participants, and/or other stakeholders during department activities to gauge how well we're doing.
37	There are adequate records of past change efforts and what happened as a result.
	Teams
38	Our department currently operates via (or is transitioning towards) a team-based structure where work projects are intentionally assigned to work groups rather than individuals with shared accountability and leadership.
39	Department employees are provided adequate training on how to work as a team member.
40	Team meetings in the department address both team processes and work content.
41	Team meetings in the department strive to include everyone's opinion.
42	Teams and work groups in the department are encouraged to learn from each other and to share their learning with others.
	Evaluation
43	The integration of evaluation activities into our department's work has enhanced (or would enhance) the quality of decision-making.
44	Managers and supervisors in the department like (or would like) staff to evaluate their efforts.

Table D1 (continued)

Question	**Evaluation (continued)**
45	Evaluation helps (or would help) the department provide better programs, processes, products and/or services.
46	There would be support among department employees if we tried to do more (or any) evaluation work.
47	Doing (more) evaluation would make it easier to convince department and organizational leadership of needed changes.
48	There are evaluation processes in place that enable department employees to review how well changes we make are working.
49	When the department engages in evaluation activities, the goal is to improve programs.
50	Data are routinely collected during department activities to inform evaluation efforts.

Note. [a]These headings represent modified six dimensions of the instrument.

Table D2

Items Associated with Institutional Demographics

Question Text	Response Scale
Which best describes the governance of your institution? Please answer according to operating authority, rather than ownership. For example, the Seattle Aquarium is owned by the city of Seattle, but operated by the non-profit Seattle Aquarium Society. This would be considered 'non-profit.'	For-profit Non-profit Public (government/municipal)
What is your annual institutional budget?	Small (< $2 million annually) Medium ($2-6.9 million) Large ($7-26 million) Extra-large (> $26 million)
What is your institution's annual attendance?	Small (< 100,000 annual visits) Medium (100,000-299,999) Large (300,000-600,000) Extra-large (> 600,000)

Table D3

Items Associated with the Work with Professional Evaluators Variable

Question Text	Response Scale
What is the name of the department/unit you oversee?	Open
What is your position title?	Open
Does your institution have dedicated internal evaluation or social science research staff? Choose all that apply, but only choose a single designation for each individual staff person. For example, if a staff person engages in both evaluation work and social science research, choose one designation or the other based on which function is more significant in their job responsibilities.	Program evaluator(s) Social science researcher(s) Audience researcher(s) No internal staff meet these criteria
How many total staff are dedicated to evaluation or social science research (FTEs)?	Open
In which department(s) do these staff work?	All these staff work in my department None of these staff work in my department Some work in my department, some work in other departments
Which other department(s) employ(s) evaluation or social science research staff?	Open
How often does your institution work with external evaluators or social science researchers? Choose all that apply. If you choose other, you will have the option to describe. □ Contract/consultants □ Audience Researchers □ University Partnerships □ Other	Weekly Monthly Several times a year Once a year Every few years N/A
If you chose 'other' in the above question, please describe	Open
Are there other staff at the institution that are comparable to professional evaluators in their knowledge and/or experience in the theory or practice of evaluation? This could include staff with formal training/education in evaluation and/or several years of work experience in an evaluative function.	Yes No Not sure

Table D3 (continued)

Question Text	Response Scale
If yes, how do they contribute to the evaluation capacity of your department or institution?	Plan, or assist other staff in planning, evaluative activities (identify outcomes, create logic models, write evaluation questions) Conduct or assist other staff in the conduct of, evaluative activities (consult on design of tools, organization or analysis of data) Consider, or work with other staff on the consideration of, the implications of evaluation findings (craft recommendations, assist with reflective activities) Other (space will be provided for details)
If you chose 'other' in the above question, please describe.	Open

Table D4

Items Associated with Phase Two Participation

Question	Question Text	Response Options
1	In phase two of this study, we will be asking a sub-set of respondents that meet a set of criteria (related to size of institution, experience with professional evaluators, and responses to this survey) to participate in an interview to better understand the context of their responses and to develop a deeper understanding of the relationship between work with professional evaluators and a department's evaluation culture. Can we contact you to request your participation in phase two? Providing your contact information now does not obligate you to participate, and you may cease participation at any time. Participants in phase two of the study will receive a summary of our discussion and a set of recommendations in addition to the summary of survey results that all phase one participants will receive.	Yes, I am open to being contacted about potential participation in phase two of this study. I understand that answering in the affirmative and providing my contact information does NOT obligate me to participate. No, please do not contact me about participation in phase two of this study.
2	Name	Open
3	Phone	Open
4	Email	Open
	Thank you for your time and expertise in completing this survey. When complete, a summary of results will be sent to the email address associated with this submission. If you have any questions, concerns, or complaints about this study, contact Jim Wharton at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.	

Appendix E: Letters of Permission

Instrument Permission

Permission was sought for the use and modification of the ROLE instrument.

Email request language.

Sent: Sunday, June 30, 2019 9:07 PM
To: Hallie Preskill [email omitted]
Subject: Permission to use the ROLE instrument

Greetings Dr. Preskill,

I have attached a letter of request to use the ROLE survey instrument in my dissertation project. Also attached you will find a brief study summary. I appreciate your consideration.

Jim Wharton
Director of Conservation Engagement and Learning
Seattle Aquarium

Letter of request language.

Greetings Dr. Preskill:

My name is Jim Wharton. I am the Director of Conservation Engagement and Learning at the Seattle Aquarium. I am also completing a Ph.D. program at the University of South Florida in educational Measurement. Members of my committee include Dr. Liliana Rodriguez-Campos, Dr. Robert Dedrick, Dr. John Ferron, and Dr. Waynne James.

I am writing to ask permission to use the Readiness for Organizational Learning and Evaluation (ROLE) survey instrument in my dissertation project. In this case, "use" means slight modification, and publication in my final dissertation. As a career-long educator and conservationist working in aquariums and science labs, I have become very interested in how our organizations build a successful culture to support our conservation missions. I am interested in understanding the relationship between an organization's work with professional evaluators and the development of their evaluation culture. I've attached a very brief summary of my proposed study, but I'd also like to explain a little why I think the ROLE instrument is a good fit for this study and how I would like to deploy it.

For the context of this study, I have developed study-specific definitions for evaluative thinking and evaluation culture (influenced by my literature review, of course). In my view:

Evaluative thinking is a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making.

An evaluation culture is one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions…with the mandate and support of organizational leadership.

While developing a new instrument is always an option, I believe the six dimensions of the ROLE instrument address the key elements of these definitions well and without modification—though I have also considered streamlining the survey by removing the Communications and Teams dimensions (as Graham and Nafukho have done in several studies) to reduce the time of completion and potentially increase my return. The only other modification I would make would be change the demographics-related questions to match my audience (the education directors at accredited US zoos and aquariums) and to add several questions to identify the extent to which these organizations (and more specifically, the education work groups within these orgs) work with professional evaluators. Once the surveys are collected, my plan is to create an overall "evaluation culture" score by summing the means of the six dimensions scores. I would use these scores to look for a relationship between the development of an organization's evaluation culture and their identified work with professional evaluators. This survey work would be followed by interviews from a sample of survey participants to better understand the context and history behind these results. We would use the survey results as the foundation of our conversation.

With your permission, I am looking forward to submitting my methodology to USF's IRB for approval this summer. I apologize for not including your co-author in this request, but I was unable to locate Dr. Torres' contact information. I would be happy to spend some time on the phone discussing this work, if you have additional questions. I would also be grateful for any references you might direct me to that utilize the ROLE instrument. I have seen your publications and three papers from Graham and Nafuhko but would be eager to learn from additional applications.

Wharton study summary language.

What is the Relationship Between Working with Professional Evaluators and an Organization's Evaluation Culture?

Zoos and aquariums are moving from tourist attractions with educational benefits to conservation organizations (Ogden & Heimlich, 2009; Brewer, 2001; Fraser & Wharton, 2007; Rabb, 2004) with aligned conservation missions against which their success or failure must be measured. With this growing emphasis, zoos and aquariums are insisting on credible metrics to help them understand how they can achieve greater mission impact. Evaluation is one tool that can provide these measures.

An organization interested in improving mission performance should strive for a culture where management supports staff who regularly engage in reflective, improvement-oriented practices which systematically collect/use data to make context-appropriate decisions, including engaging in formal evaluation activities as warranted. This can be described as a strong *evaluation culture*. One way to develop evaluative thinking or an evaluation culture with staff is to work with evaluators. Increasingly, these "process" benefits are valued as highly as the instrumental benefits provided by the findings of any individual evaluation (Hargreaves & Podems, 2012; Patton, 1998; Preskill & Zuckerman, 2003).

Some organizations have invested in internal evaluation capacity (one or more staff whose primary responsibility is evaluation and who possess some professional experience or training), while others work exclusively with external evaluators. Still others have chosen not to work with professional evaluators in any capacity.

Understanding the relationship between the manner in which an organization works with professional evaluators and its evaluation culture (and thereby its mission performance) is important, because this investment represents a strategic use of limited resources that could be employed for program development or implementation.

I am interested in surveying the leadership from programming departments at accredited U.S. zoos and aquariums to learn how working with professional evaluation staff is related to the nature of the evaluation culture at these mission-based organizations. The study will be conducted in two phases. First will be a survey of education directors at AZA institutions using a modified version of the Readiness for Learning and Evaluation (ROLE) instrument (Preskill & Torres, 2000), which maps closely to this study's definitions of evaluative thinking and evaluation culture. The ROLE instrument generates a score for evaluation culture which will be used to divide responses into strong, moderate, and weak categories. The second phase will consist of follow-up interviews with education directors from each category to better understand how their scores might relate to their experience working with professional evaluators.

I expect to find that institutions which have invested in internal evaluation staff will be associated with the strongest evaluation cultures (as assessed in this study) and that organizations that work with a combination of internal and external evaluators may be associated with slightly stronger evaluation cultures, but the differences will not be significant in survey results.

Using evaluation can show an institution how their programs might be improved, but how do we know if investments in evaluation improve our organizational culture and/or overall mission performance? Studies like this one can begin to answer this question by establishing a relationship between this work and the strength of an evaluation culture. Further studies will be necessary to understand the mechanisms through which working with professional evaluators positively effects the development and maintenance of a such a culture.

Response to request.

Sent: Monday, July 1, 2019 5:11 AM
To: Jim Wharton [email omitted]
Subject: RE: Permission to use the ROLE instrument

Hello Jim,

Thanks for reaching out – your study looks really interesting!

Yes, you have my permission to use the ROLE survey – since it's in the public domain, you really don't need my permission. I wish you all the best with your dissertation research and its completion.

Best regards,

Hallie Preskill, PhD.
Managing Director
FSG.org

Letter of Support from the Association of Zoos and Aquariums

Email request language.

Sent: Sunday, January 26, 2020 4:43 PM
To: Amy Rutherford [email omitted]
Subject: Letter of Support for Dissertation Work

Hi Amy,

I know we've talked a little about my dissertation project, but I can't remember if we discussed this bit or not (mostly because it's taking me so long to get anything done 🙄).

My project is interested in how working with professional evaluators influences the evaluation culture of a workgroup/institution. I've attached a short study summary.

Phase one of the project involves a short survey that is directed at the leading managers (directors) of the programming departments at AZA-accredited zoos and aquariums. I would like to use the AZA-education contact list to send an invitation for participation (also attached). To do this, IRB is requesting a letter of support from you to confirm my access to this list.

Is this something you could provide? Of course, I could construct this list independently from my own contacts, the AZA Network, and an internet search, but that is A LOT of (sort of phony) work when this list is available. Posting a hopeful, general call on the Network and associated listservs is unlikely to generate a response comparable to a personalized direct invitation. I am not asking AZA to endorse the study by providing this list...though if that were possible, it would certainly bolster the case for participation. I plan to share the results of the study with the AZA community including CEC and the Midyear and Annual meeting audiences, possibly Connect, if there was interest.

Happy to answer any questions. Let me know what you think.

JMW

Response to request.

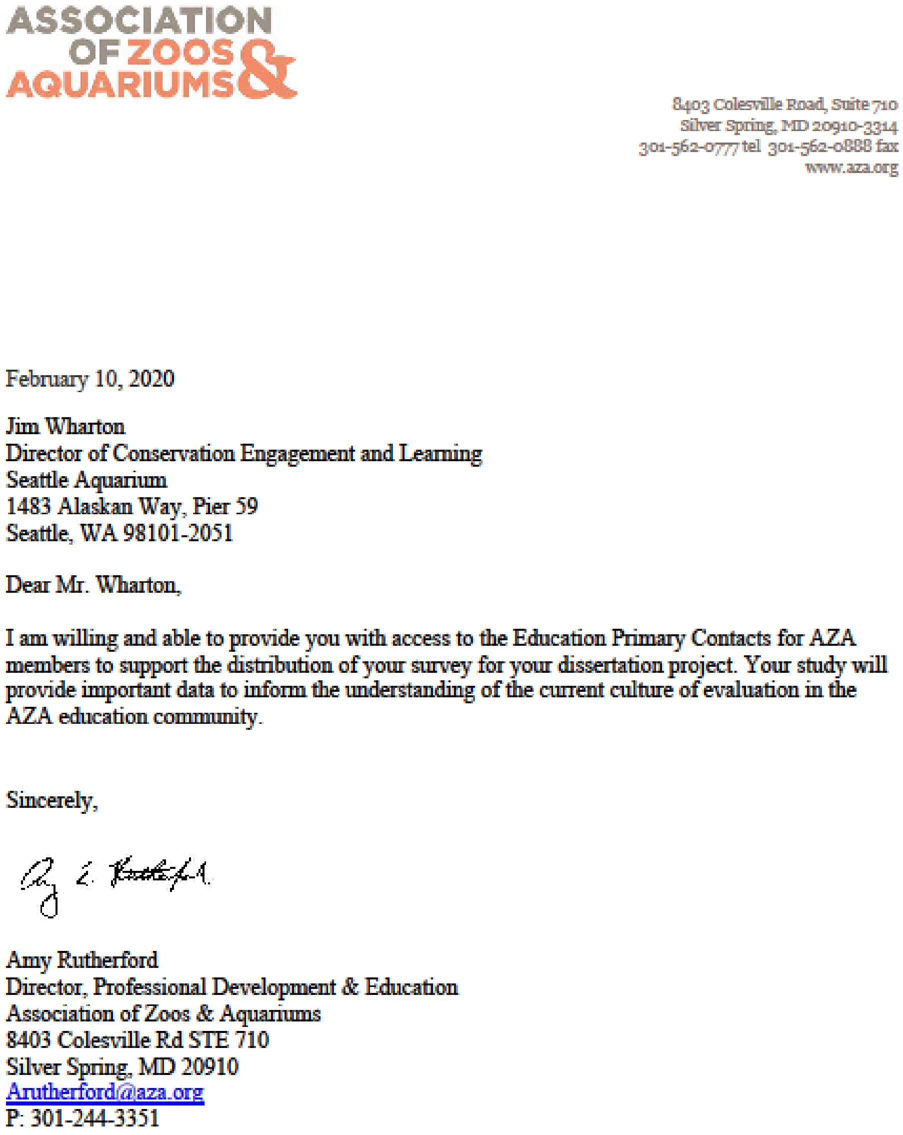
Sent: Monday, February 10, 2020 6:48 AM
To: Jim Wharton [email omitted]
Subject: RE: Letter of Support + Travel for March Meeting

Hi Jim,

Attached please find the letter of support for your dissertation survey. Let me know your timeline on getting the survey out and I can pull the most up to date list for distribution at that time.

Amy Rutherford
Director of Professional Development & Education
Association of Zoos & Aquariums

Content of AZA letter of support.



ASSOCIATION OF ZOOS & AQUARIUMS

8403 Colesville Road, Suite 710
Silver Spring, MD 20910-3314
301-562-0777 tel 301-562-0888 fax
www.aza.org

February 10, 2020

Jim Wharton
Director of Conservation Engagement and Learning
Seattle Aquarium
1483 Alaskan Way, Pier 59
Seattle, WA 98101-2051

Dear Mr. Wharton,

I am willing and able to provide you with access to the Education Primary Contacts for AZA members to support the distribution of your survey for your dissertation project. Your study will provide important data to inform the understanding of the current culture of evaluation in the AZA education community.

Sincerely,

Amy Rutherford
Director, Professional Development & Education
Association of Zoos & Aquariums
8403 Colesville Rd STE 710
Silver Spring, MD 20910
Arutherford@aza.org
P: 301-244-3351

Appendix F: Survey Instrument Review

Letters of Request to Reviewers

Survey instrument. Sample letter or request sent to reviewer.

To: [reviewer]
Subject: Help Jim Wharton with his dissertation

Hi [reviewer],

Hope you are holding up alright through all of this history we're living through. So [you may know/I don't think we've chatted about this before but] I've been working on my PhD in educational measurement. I'm finally through IRB and ready to get this project moving. Because of your experience with our field, I was hoping you would be willing to lend your expert eye to my survey instrument. I've linked the instrument I'm planning to use in the study. It is a largely intact version of the Readiness for Organizational Learning and Evaluation (ROLE) instrument from Preskill and Torres (2000). I'm attaching a study summary here (Wharton study summary v.3), but all the information a reviewer should need is in the 'reviewers version' of the survey linked below.

The survey is estimated to take 25 minutes or less to complete. There are additional questions added to facilitate review. The whole effort could take less than 45 minutes (or longer, of course, depending on how much you have to say).

If this isn't something you have time or interest in, just let me know. I appreciate any help you can offer.

Thanks in advance,

JMW

Follow this link to the Survey: [Take the Survey]

Follow-up letter to reviewers.

To: [reviewer]
Subject: Near Final Survey

Hi [reviewer],

Thanks for your review of the initial instrument. I have made many changes and wanted to share the penultimate version with you. A summary of changes:

- I significantly shortened the survey, removing (on balance) 24 items.
- I removed the categories and definitions from the introduction and survey format to minimize cuing.
- I added specific language to each item so that there would be no confusion about the context of the item.
- I added all 50 questions in a single randomized block to further reduce cuing and inserted page breaks to minimize scrolling.
- I added a progress bar and back buttons.
- I switched from discrete Likert response options to a bar with a greater response range for better item variance and reliability. (All responses are now on the same scale...no out-of-place binary response sets)
- I removed the 60% qualifier for evaluation staff.
- I corrected several structural issues and typos (including one that required respondents to agree to the second phase before proceeding).
- There are still areas where more depth or specificity might be helpful, but I am counting on the interview phase to provide these important details.

Thanks again for your feedback. Right now, I am sharing the survey with my department leadership team for one last usability review, then plan to distribute in July. If you had any last comments or suggestions, of course, they'd be welcome. Onward!

[Link to survey]

Interview protocol. I asked three of the five survey reviewers to review the interview protocol. The initial contact was done informally (via text or phone call).

To: [reviewer]
Subject: Near Final Survey

Hi [reviewer],

Thanks for agreeing to take a look at this. I am attaching both the script and my study summary to refresh you on the project. Generally, there will be 2-3 interviews from high/med/low scoring institutions on the evaluation culture measure. They'll be recorded Zoom interviews.

I'm in the midst of dealing with data analysis, so not in a huge rush. Any time in the next few weeks is fine.

Thanks again,

JMW

Reviewer Questions

The following questions and language were added to the reviewers' version of the survey to facilitate feedback.

Reviewer Introduction. Thanks for agreeing to review this study instrument. In addition to the information in the study summary, here is some additional information to help in your evaluation. I am also happy to share my approved IRB protocol or dissertation proposal, if either would be helpful.

There are two constructs addressed by this instrument, evaluation culture and evaluative thinking.

For the purposes of this study, evaluative thinking is defined as a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making. The definition is similar to Baker and Bruner (2006) with the addition of social learning, the identification of assumptions and positionality, and the idea of organizational context appropriateness.

Likewise, in the context of this study, an evaluation culture is one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions—with the mandate and support of organizational leadership.

The bulk of this instrument is an unmodified version of the Readiness for Organizational Learning and Evaluation (ROLE) survey (Preskill & Torres, 2000). The ROLE survey has six dimensions that map closely to the study's definition of evaluative thinking: culture, leadership, systems and structure, communication, teamwork, and evaluation. The balance of the survey asks about the details of the institution's characteristics (size, operating authority, etc.) and work with evaluators (the study's independent variable).

What I am hoping for from your review is your professional opinion on survey design, face and content validity, and any other feedback you might consider helpful. The wording of items associated with the ROLE instrument are fairly fixed, but the presentation of those items could be done in a variety of ways. In this instrument, I have broken them up as separate "questions." This is also my first project using Qualtrics, so any tips or tricks you might like to share would be valuable to me.

If you have questions or clarifications before or during your review, you can contact me at jmwharto@mail.usf.edu or [phone number]. At the end of this survey there will be several Likert-style questions and corresponding open text boxes for you to provide your feedback. If you would prefer to share your thoughts over the phone or in a separate email, that is also just fine.

Table F1

Items Added to Reviewers' Copy to Solicit Feedback

Question	Question Text	Response Scale
1	How do you feel about the length of the survey?	1 (Too Short) 2 3 (Just Right) 4 5 (Too Long)
2	What other comments do you have about the survey design?	Open
3	How do you feel about the face and content validity of the survey? For your reference, here are the two relevant definitions again: Evaluative thinking is defined as a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making. The definition is similar to Baker and Bruner (2006) with the addition of social learning, the identification of assumptions and positionality, and the idea of organizational context appropriateness. An evaluation culture is one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions—with the mandate and support of organizational leadership. ☐ The survey adequately addresses the study definition of evaluative thinking. ☐ The survey adequately addresses the study definition of an evaluation culture. ☐ The study asks the most relevant questions about institutional and work group characteristics. ☐ The survey asks the right questions about work group interactions with professional evaluators.	Likert 1-5 Strongly disagree- Strongly agree
4	Do you have additional feedback about how the survey does or does not address the constructs of evaluative thinking and/or evaluation culture within the context of this study?	Open
5	Do you have additional feedback about the information collected by the survey about the institution and work group?	Open
6	Do you have additional feedback about the information collected by the survey about the work group's interactions with professional evaluators?	Open
7	Do you have other comments or suggestions?	Open

Table F2

Summary of Reviewer Feedback on Survey Instrument

Recommendation/Feedback	Changes Made
Tough time shifting from the evaluation context to the broader work context (especially Teamwork and Leadership sections). Would some additional framing or reminders help? Subheadings for sections is suggested.	Removed sections, changed 'employees' with 'department employees', adding "in your department" to some items to help with framing.
Why are the first three Teams questions binary? Do they have to be?	Changing all questions to same format.
Survey isn't sufficient to get at the nuances of evaluative thinking, especially the social aspects. Kept wanting to go deeper.	None. Interviews in phase two is where this will develop.
May need to add more language to distinguish between program evaluation and visitor studies. Some institutions ONLY do visitor studies and therefore the 'evaluators' live in the marketing or communications department.	None.
Asked about what *kind* of evaluation (outcomes based, satisfaction based, etc). Would an org with a well-developed practice of satisfaction-based evaluations score highly as a 'strong' evaluation culture?	No change. Will see if anything notable emerges of concern.
Change PE5 from "Does your institution..." to "How often does your institution..."	OK
PE5 Is there really no option to share what 'other' is?	Language updated to make skip logic clearer.
"There's also a question in this section (PE7-9) about whether you have staff who spend less than 60% of their time on evaluation. The next questions pigeonhole the responses into staff who support other staff to do evaluation, not the program staff themselves (for example, we had trained all program staff - educators, coordinators, managers - to carry out evaluations in conjunction with our dedicated internal evaluation staff. But it was not their full-time job). So the wording of the responses was a bit limiting."	Changed language, removed open-ended question and added language to include doing, in addition to assisting in, evaluation activities. Other an additional option.
***Re-read for misspellings that would be missed by spell check. (Intro: ...study does note obligate...")	Done.
Suggest using page headers to help with framing.	No longer nec. With removal of dimensions and item randomization this becomes moot.
Suggest adding progress bar...debate about utility, but she wondered a couple of times if the survey would go "on and on."	OK

Table F2 (continued)

Recommendation/Feedback	Changes Made
Consider a back button. Found herself wanting to review what she'd previously said.	OK
Ask how many staff (department size) + how many staff dedicated to evaluation (>60%).	Added.
Why use 'somewhat'? Is this what the ROLE uses? Why not just say 'agree' or 'disagree'?	Moot point as we are switching to VAS bars.
Strongly agree on face and content validity.	
Made a note about a question from the evaluation section (about starting/doing evaluation) in the field about the meta questions concerning working with professional evaluators.	
Would have preferred to comment on each block of questions right after taking them.	
Found survey to be a little long.	Shortened substantially.
Felt survey addressed study definitions (SA).	
Fine with info collected about institution.	
Double-check on formatting on IRB language. It is appearing in different colors and fonts to some. Use Rich content editor to fix.	I scrubbed the formatting and started over. Doesn't look beautiful, but the editor is clunky.
Does the language about interview in the IRB section intimidating? Will some opt out because they feel like they HAVE to participate in the interview?	Re-read, but left unchanged. Very little leeway to deviate from the approved language.
Is evaluation thinking too jargony for non-evaluators? Maybe add examples. Whole page is text heavy.	Removed definition.
Will providing definitions create a response bias?	Killed these to avoid this and to further reduce wordiness of the instrument.
If you add a definition for evaluation for the evaluation questions, shouldn't it be included in the beginning with the other definitions?	Pulling with elimination of dimension categories and item randomization.
May need an example to help people understand the 60% qualifier for evaluation, maybe also for other evaluator questions	60% rule eliminated altogether.
Should there be an I don't know option for the question about having knowledgeable evaluation staff?	Added.
Fairly text heavy throughout. Could use a pass to make it leaner.	Cut introduction, edited throughout.
Break up question sets even more for mobile users? Tradeoffs because then you're click next A LOT.	

Table F2 (continued)

Recommendation/Feedback	Changes Made
More instructions for how to answer?	Likely unnecessary with shortening and simplification of structure.
Too many ANDs in questions.	Made changes where seemed problematic, but several instances seemed like they were providing examples of the construct interrogated rather than introducing two different constructs.
5-point scale vs slide?	Understand rationale. Sliders, though, have mixed support in the literature. VAS bars seem more equivalent to radio buttons. Will switch to these. May need some "IDK" options.
Suggests doing sections "as an employee" "as a supervisor" and "in my role in the org."	Will add specifying language to questions that won't change the nature of the question but will clarify the frame.
Evaluator questions need to be more nuanced (?). Not just what they do together, but how they do it. Example given for perfunctory grant evaluation.	If an org works regularly with PEs but does bad work...they should score low. If they score high...but the work is bad, then maybe there is a problem...but if they score low and the work is bad, that fits the model. These are things to work out in the interviews once we see the results and look at the trends in the data. I agree, but I'm not sure the survey is the place to get at those questions.
Reverse wording questions.	Committee member advises against, finds it can be confusing for respondents.
Suggest replacing "supervisors and managers" with the generic "managers" or "leadership"	Changed instead to "managers and supervisors in the department."
Asking questions and raising issues are two separate constructs.	Disagree. The intent of the question is whether staff feel comfortable speaking up to management.
Employees are encouraged to take the lead in initiating change OR trying to do something different. (Two separate constructs).	Same as above. Examples of employees feeling free to change direction rather than accepting the ways things are being done.
"Employees are available..." ...aren't relevant to an evaluative culture. Being present physically is not necessarily being open to deliberative reasoning. You need to do something that's more about the value of how they interact and contribute."	Item removed in pruning.
Switch from "employees" to I statements (see examples from email).	Switched to "department employees" because this is an assessment of how they view their department as working, not how they feel about their work group.

Table F2 (continued)

Recommendation/Feedback	Changes Made
"Employees meet deadlines" is just a bad question.	Question eliminated in pruning for length
Lumping too many stakeholders in a single question.	The questions is about whether information is generally gathered...again, these are just examples to help with their thinking.
Binary Teams questions are inadequate. See email for suggestions.	All are moving to same format.
Doesn't understand the 60% rule.	Killing this. It was set by me to indicate evaluation was their primary function but may be over-thinking it. If they are called evaluator, then they are an evaluator.
Can you freeze the top row in the Likert?	Will, if format allows.
need to be clear about which team...their department or their executive team (to which they may belong).	Added language for specification of framing to many questions.
Confused about section on supervisors and managers when the survey takers ARE the supervisors and managers.	Language clarification made.
600,000 and up for visitor number. This category may have a much wider range than the others.	Perhaps, but it does represent a quartile. And does the range represent a meaningful difference in operation? Some, sure. We are at the low end of that range, Georgia and San Diego at the top. Are we similar orgs?
Need to keep moving, starting over with a "better" instrument not advised.	
Could split items if AND is a concern.	Reviewed language and modified in a few cases
Sliders are fine, either or.	Sliders have poor support in literature, VAS bars better supported and give the increased variance desired.
Eliminating items is not a problem for IRB, Dedrick supports shorter instrument for increased response rates.	Significantly shortened the instrument by 30 items (but added two to supplement evaluation and data collection.
Some questions seemed repetitive.	No change
Length appropriate.	
100-point scale requires more thought.	Could length time for completion but should create more variability.
Why only supervisors? Should statements say 'I believe' since they are answering for others?	Added "In some cases, you will be asked to judge the opinions or feelings of your department employees to the best of your ability." to be clear about what is expected.

Table F2 (continued)

Recommendation/Feedback	Changes Made
Review for we/I vs they/them.	Ended up not to be an issue.
What are 'work issues'? Will this be confusing to people?	Decided no change on this.
Check for accidental cut/paste repetition errors	Found it.
Look at spacing on "Info is gathered question."	Reviewed.
Look at 'Doing more evaluation' wording...	Not changing. Maybe not perfect wording but may be clearer from other answers.
Look at "team based" question for wording.	Reviewed.
Is budget pre-COVID?	Added: "In a typical, non-COVID year" to budget and attendance questions.
Switch departments to department(s)	Done.
Review format of initial consent questions.	~~Added: Click or touch statements to indicate your agreement"~~ Format changed when I switched to the generic USF template, so removed.
Should there be an 'I don't know' option?	Decided against.
Demo questions much improved.	
See if you can drop the USF "Health" from the survey format.	Found this and switched formats.
Review for number agreement on data.	Found it.
Will participants understand 'Managers and supervisors'?	This is common language in the field.
Change "directors" to you to be clearer I am talking to them, not their bosses.	Reviewed.
Clarify that I am asking them to answer as...	Reviewed.
Who do employees talk to about work issues, share their opinions?	Question eliminated in pruning.
Teams-based structure...definition needed?	Added a definition that specified shared accountability and leaderships.

Appendix G: Survey Correspondence

Phase One and Two Correspondence

What follows is the email language used to solicit participation in phases one and two of the project.

Pre-invitation letter.

Sent: Monday, July 27, 2020 2:12 AM
To: [Participant]
Subject: Watch for an invitation this week to learn more about your institution's evaluation culture

Good morning [Participant],

On Wednesday of this week, you will receive a link to a survey I am conducting as part of my doctoral dissertation. I am surveying the education/engagement directors at US-based, AZA-accredited institutions to learn more about how our work with professional evaluators might influence our departments' and institutions' evaluation cultures. A smaller group of volunteer respondents will be invited to participate in a round of interviews in phase two of this study.

I invite you to set aside 15-20 minutes this week or next to complete the survey. If you are not the senior manager of the education/engagement department at your facility, I would appreciate it if you could forward the forthcoming invitation to the appropriate party. If you would like me to change the invitation address for your institution, feel free to send me the new information and I will make sure the survey invite gets sent directly to them. I know right now staff and hours are in flux, so I appreciate your help with this.

Participation in this study is voluntary and confidential. You may cease participation at any time. You will not be compensated for participation, but you may find a summary of your responses in the context of the study sample valuable to you personally or professionally. Any data or findings shared as a result of this study will be anonymized. The study's procedures have been reviewed and approved by the Institutional Review Board at the University of South Florida.

If you have any questions, concerns, or complaints about this study, you may contact me at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

I appreciate and look forward to your institution's participation. Understanding how our investment in professional evaluation services may contribute to the organizational culture of our institutions could return significant benefits to our effectiveness as conservation organizations.

Thanks in advance for your help,

Jim Wharton
Director of Conservation Engagement & Learning
Seattle Aquarium

Invitation to participate in survey.

Sent: Wednesday, July 29, 2020 8:02 AM
To: [Participant]
Subject: This is it: Improve your department's evaluation culture and help me with my dissertation!

Good Morning [Participant],

I am reaching out to you today for your help on a project that I believe could help AZA institutions improve the professional cultures of our organizations to better serve our missions. I'm conducting a study as part of my doctoral dissertation looking at the relationship between our work with professional evaluators and our departments' and institutions' evaluation cultures. You've been included because AZA has identified you as the senior manager of the education/engagement function at your institution. However, it's also possible we didn't have the current information for that position and have included you as an active contact in the leadership of your institution. I would appreciate it if you could forward this invitation to the appropriate party. These are crazy times with staff and hours in flux. Thanks in advance for your attention to this.

As AZA-accredited facilities, our institutions are required to evaluate our conservation and education efforts as part of the accreditation process. As an industry we invest millions in evaluation, not just to meet these standards, but to improve our ability to achieve our conservation missions. How these efforts are affecting our professional culture is an important question to answer, especially now as a global pandemic puts more pressure on our institutional bottom lines than ever before.

I would sincerely appreciate 15-20 minutes of your time in the next two weeks to complete this survey. Please see the postscript of this message for important information about your participation. A small subset of volunteer respondents will be invited to participate in a round of interviews in phase two of the study. This is not required for participation. Your responses to the survey are very valuable in and of themselves.

Follow this link to the Survey: [Take the Survey]
Or copy and paste the URL below into your internet browser: [URL]

Thank you in advance for your time and expertise.

Jim Wharton
Director of Conservation Engagement & Learning
Seattle Aquarium

P.S. Participation in this study is voluntary and confidential. You may cease participation at any time. You will not be compensated for participation, but you may find a summary of your responses in the context of the study sample valuable to you personally or professionally. Any data or findings shared as a result of this study will be anonymized. The study's procedures have been reviewed and approved by the Institutional Review Board at the University of South Florida.

If you have any questions, concerns, or complaints about this study, you may contact me at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

AZA Education Listserv communication. To try to catch any surveys that landed in the wrong inbox, or that might have gotten caught in spam filters, I also send a notification to AZA's education listserv. This is populated by education staff at AZA institutions and resulted in several follow-up messages.

Sent: 07-30-2020 15:02
From: Jim Wharton
Subject: Survey on evaluation and our evaluation culture...

Hi List,

Yesterday I sent a survey invitation to the education leadership at each U.S.-based AZA facility. The survey is part of my dissertation work and addresses how our work with professional evaluators (internal and external) might relate to the evaluation culture of our organizations. If you are the top education manager at your facility and you have NOT seen this invitation, please check your spam folder for it. If it is totally missing, please drop me a line. Things being as they are, the status of education contacts at each facility is very much in flux. Yes, this is a pretty crappy time to try to finish this project, but life and work must go on, right?

If you've received the invitation and have 15-20 minutes to complete it, I would be eternally grateful...and you may even find the results interesting and constructive for your department's work and culture. If you are not the top education manager, maybe gentle nudge for your boss?

Thanks all. I'm excited to share the results when they are complete. Feel free to contact me if you have questions.

Jim Wharton
Director of Conservation Engagement and Learning
Seattle Aquarium

Letters to institutions with email errors.

Sent: Thursday, July 30, 2020 1:50 AM
To: [Participant]
Subject: An opportunity to learn more about your AZA institution's evaluation culture

Good Morning [Participant],

My name is Jim Wharton. I am the Director of Conservation Engagement and Learning at the Seattle Aquarium. Yesterday I sent an invitation to the AZA education contact at your institution, but it bounced. I know that there has been a lot of turmoil and flux in staff and hours during these complex times so I wanted to reach out to you to make sure your institution had an opportunity to participate in a project that I believe could help our organizations improve the professional cultures of our organizations to better serve our missions.

I'm conducting a study as part of my doctoral dissertation looking at the relationship between our work with professional evaluators and our departments' and institutions' evaluation cultures. As AZA-accredited facilities, our institutions are required to evaluate our conservation and education efforts as part of the accreditation process. As an industry we invest millions in evaluation, not just to meet these standards, but to improve our ability to achieve our conservation missions. How these efforts are affecting our professional culture is an important question to answer, especially now as a global pandemic puts more pressure on our institutional bottom lines than ever before.

I'm hoping you can forward this survey to the top education manager at your facility. The survey should take just 15-20 minutes of their time. Please see the postscript of this message for important information about their participation. A small subset of volunteer respondents will be invited to participate in a round of interviews in phase two of the study. This is not required for participation. Your institution's responses to the survey are very valuable in and of themselves.

Follow this link to the Survey: [Take the Survey]
Or copy and paste the URL below into your internet browser: [URL]

Thank you in advance for your time and expertise.

Jim Wharton
Director of Conservation Engagement & Learning
Seattle Aquarium

P.S. Participation in this study is voluntary and confidential. You may cease participation at any time. You will not be compensated for participation, but you may find a summary of your responses in the context of the study sample valuable to you personally or professionally. Any data or findings shared as a result of this study will be anonymized. The study's procedures have been reviewed and approved by the Institutional Review Board at the University of South Florida.

If you have any questions, concerns, or complaints about this study, you may contact me at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

Survey follow-ups. Two follow-ups were sent to participants who had not completed the survey.

First follow-up.

Sent: Wednesday, August 5, 2020 8:01 AM
To: [Participant]
Subject: A gentle reminder to fill out your survey about your department's evaluation culture

Good morning [Participant],

Last Wednesday you received an invitation to participate in a project assessing the relationship between our work with professional evaluators and our institution's evaluation culture. If you haven't had 15-20 minutes yet to dedicate to the survey, this is a gentle reminder that the survey will remain open for one more week, closing on August 12. Your institution's participation would be deeply appreciated, and I hope seeing the results from across our industry will be valuable to you.

Follow this link to the Survey: [Take the Survey]
Or copy and paste the URL below into your internet browser: [URL]

If you have any questions, concerns, or complaints about this study, you may contact me at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

Jim Wharton
Director of Conservation Engagement & Learning
Seattle Aquarium

Second follow-up with deadline extension.

Sent: Wednesday, August 12, 2020 8:01 AM
To: [Participant]
Subject: Survey deadline extended: Still time to learn more about your evaluation culture and help Jim Wharton with his dissertation!

Good morning [Participant],

I hope this email finds you well with [Participating Institution] open and thriving safely. I've decided to extend the survey deadline one more week as these are crazy times and I want to make sure everyone who would like to participate has the opportunity. As a reminder, this is a project designed to explore the relationship between our work with professional evaluators and our organization's evaluation cultures. You can find the survey at the following link through August 19. It's intended for the top education/engagement manager at your facility.

Follow this link to the Survey: [Take the Survey]
Or copy and paste the URL below into your internet browser: [URL]

If you have any questions, concerns, or complaints about this study, you may contact me at [phone number] or [email address]. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

Thanks again for your consideration,

Jim Wharton
Director of Conservation Engagement & Learning
Seattle Aquarium

Final request for those that started but did not complete surveys.

Sent: Monday, August 17, 2020 1:11 AM
To: [Participant]
Subject: Thanks for starting my survey. Be sure to complete by August 19.

Good morning [Participant],

Thanks for starting my survey. I just wanted to make sure you knew the survey will close this Wednesday, August 19 at midnight PST. If you had intended to get back to it, this is a gentle nudge to reopen and finish it off. If you'd opened it and passed it along or determined it wasn't something you had time to complete, thank you for reviewing it.

I sincerely appreciate your time,

Jim Wharton
Seattle Aquarium

Phase two interview invitation.

Sent: Monday, November 23, 2020 7:31 AM
To: [Participant]
Subject: Willing to participate in an interview with Jim Wharton?

Hi [Participant],

Thanks for participating in phase one of my dissertation research on the relationship between work with professional evaluators and the evaluation cultures in our institutions. Attached is a summary of your responses compared to the sample of 100 education directors drawn from AZA-accredited zoos and aquariums around the country.

I'm hoping you will be willing to be one of a small group of follow-up interviews in phase two of the project. I would like to learn more about the evaluation culture at your organization. I divided the sample into three tiers based on their overall 'evaluation culture' score (the average of the six means from the survey instrument's subscales), and I am choosing three institutions from each tier. Your scores placed the [Participating Institution] in the [upper/middle/lower] tier. I am interested adding the [Institution] as an example of a [organizational circumstance].

I would like to schedule interviews in the month of December at your convenience. They will be conducted via Zoom and I will keep them to one-hour or less. You have the option of also providing supplementary documents, if you think they would be helpful to review before our interview. Participation in an interview is, of course, completely voluntary and you can pull out at any time for any reason. Attached is a copy of the consent language you completed as part of the online survey that includes contact information for the University of South Florida's Institutional Review Board (IRB) if you have any questions or concerns.

I hope you and your circle are safe and healthy. If you are willing to participate, just response to this email and I will send you a poll with scheduling options. If you would prefer not to participate, that's no problem at all, but a quick note would be helpful to let me know I need to find an alternate.

Hope you have a safe and relaxing Thanksgiving break,

Jim Wharton
Director of Conservation Engagement and Learning
Seattle Aquarium

Response Summary Template. Included in Appendix H.

Appendix H: Final Interview Protocol

Introduction

1. Thanks for participating in this study. I have number of questions to discuss with you. I'm hoping it will take us about an hour, but we could go a little long if there is a lot to talk about. Does this timeframe sound okay with you? Remember, you are welcome to stop at any time. I appreciate and value the time you are dedicating to this work and if, at any time, you would like to pause or stop the interview, this is no problem and fully your right to do so.
2. I would like to use the recording function of the video conference platform to assist in my notetaking and summary. Is that alright with you? I will also be using the screen-sharing function to present some materials for us to refer and react to.
3. Before we start, do you have any questions about the study overall and your participation? This could include questions about how your data and identity is handled or anything else you're curious about.
4. A note on terminology. I will be using workgroup and department a little interchangeably. I know AZA org charts can be complex. Could you quickly clarify how your position falls in the zoo/aquarium's hierarchy? And could you tell be a little about your department?

Working with Professional Evaluators

5. The overall goal of the study is to begin to understand the relationship between the way we work with professional evaluators and the evaluation culture (EC) of an organization (specifically the programming departments of accredited zoos and aquariums). On the screen, I'm sharing the definition of a "professional evaluator" used in the context of this study: *professional evaluators are those with formal training/education in evaluation and/or several years of work experience in an evaluative function.* Do you have any questions or clarifications regarding the study definition?
6. I want to talk a little about how you and your workgroup have worked with professional evaluators in the past. In the phase one survey, you indicated [INSERT FROM RESULTS], and in the documents you provided, there was also mention of [INSERT FROM RESULTS]. [SHOW SUMMARY ON SCREEN] Do these look accurate?
7. [IF APPROPRIATE] What more can you tell me about the [EVALUATION] staff at your institution? [POSSIBLE FOLLOW-UP QUESTIONS] How long has the team had the position? How long has the current staff worked here? Did you lose staff/do you think you will regain staff from COVID reductions? What do they work on/how do they divide their time?
8. Can you tell me a little about your evaluation background? Have you had any education at university, professional training, or on-the-job experience?
9. [IF APPROPRIATE] What more can you tell me about the [EXTERNAL EVALUATION] colleagues you have worked with? [POSSIBLE FOLLOW-UP

QUESTIONS] Clarify how often/how extensive? Do you have some you work with regularly/have a relationship with? How have they influenced the ongoing work or culture of your team?

10. [IF APPROPRIATE] What more can you tell me about the [OTHER STAFF WITH EVALUTION EXPERIENCE] staff at your institution? [POSSIBLE FOLLOW-UP QUESTIONS] Details on their training? How much do they do/lead/consult on evaluation work that is not directly related to their programmatic work?
11. How has working with these evaluators lead to differences in the way you use or view evaluation in your workgroup or at the zoo/aquarium, if at all?
12. In your opinion, has working with these evaluators influenced the way leadership views the priority, use, or value of evaluation? [POSSIBLE FOLLOW-UP QUESTIONS] In what ways?

Evaluation Culture

13. When you hear the term "evaluation culture," what does that mean to you? How would you define an evaluation culture?
14. [I am hearing a lot of things in common with/Interesting, that is quite different from] the definition I have developed for the context of the study [SHOW ON SCREEN]: *An evaluation culture is one where staff regularly use evaluation as a tool to improve programs and evaluative thinking every day to make better decisions—with the mandate and support of organizational leadership.*
15. [IF VERY DIFFERENT] Let's talk about some of the differences between the definitions. Tell me more about [DIFFERENCES]… [REPEAT AS NECESSARY]

16. [IF SIMILAR] Do you have any questions or clarifications about the study definition?
17. Thinking about this definition and your own, how would you characterize the ‘evaluation culture’ of your department?
18. Do you think the EC of your department differs from the EC of the organization? [IF YES] In what ways? [FOLLOW UP OR IF NO] Why do you think that is?
19. Your responses to the phase one survey show [INSERT INFORMATION ON OVERALL SCORE] with strengths in the areas of [INSERT HIGH SCORING DIMENSIONS] and opportunities for improvement in the areas of [INSERT LOWER SCORING DIMENSIONS]. Do these scores match the way you think about the EC of your department? **Keep in mind that these scores are not an objective assessment of these areas, but rather your contextualized judgement. You’re comparing them to other director’s contextualized judgements.**
20. [IF NO] What turned out differently? Why do you think that is?
21. Can you share any additional context around why you scored your work group so highly in the areas of [INSERT HIGH SCORING DIMENSIONS]? What about for some of the lower scoring areas like [INSERT LOWER SCORING DIMENSIONS]?
22. Do you think your staff would score the department differently? [IF YES] How do you think their scores would differ? Why do you think they would view the department differently in these areas?

Evaluative Thinking

23. A core element of the study's definition and idea of an EC is the ability of staff to think 'evaluatively.' When you hear a term like evaluative thinking, what does it mean for you? What does it mean for a staff person to think evaluatively?
24. [I hear a lot of similarities/The definition used in the study has some different elements]. For this study, I developed a definition influenced by others I found in the literature, especially Baker and Bruner (2006). Here it is [SHOW ON SCREEN]: *Evaluative thinking is a social, reflective practice woven into the everyday practices of an organization that identifies assumptions and positionality and uses systematically collected evidence to inform context-appropriate decision-making.*
25. [IF VERY DIFFERENT] Let's talk about some of the differences between the definitions. Tell me more about [DIFFERENCES]… [REPEAT AS NECESSARY]
26. [IF SIMILAR] Do you have any questions or clarifications about the study definition?
27. The key elements in there include social learning, reflective practice, examining assumptions, and data-informed decision-making. How much do these practices influence you and your staff's day-to-day work? Can you give some examples?
28. How do you think your workgroup compares to other workgroups within the zoo/aquarium? What makes you say that?
29. Have you done any professional development for your staff around any of these skills?

30. How has working with professional evaluators influenced your staff's ability to think evaluatively, if at all? Have they contributed to developing any of the elements we discussed earlier [LIST ON SCREEN AGAIN]?

Psychological Safety

31. So, you read your results summary, but no doubt noticed there were no conclusions included regarding the influence of evaluators on the evaluation culture scores. As it turns out, the survey results did not show differences in scores that could be connected to either institutional demographics or work with professional evaluations. That certainly surprised me. When I did an exploratory factor analysis, I found that the responses in this data set did not generate factors that neatly matched the subscales on the instrument we used. They did however line-up with two constructs of interest to the study (evaluative thinking and evaluation) and one emergent construct I'd like to chat about with you: the idea of psychological safety. When I use that term…what does it bring up for you? What does it mean for a workplace to have psychological safety?
32. A definition that is commonly used in the academic literature is "a shared belief held by members of a team that the team is safe for interpersonal risk-taking" (Edmondson, 1999). The items and constructs linked to the idea included [SHOW ON SCREEN]:
    - Openness to differences (in people and ideas).
    - Risk-taking.
    - Openness of leadership to for input from staff.

- Collaborative spirit.
- Employees feeling valued for what they bring to the table.
- Freedom to make and learn from mistakes.

When I pulled together the items identified by the factor analysis, here is how you scored your team in in these emergent factors [SHOW ON SCREEN]. Do these scores match how you view your work group? Is psychological safety an area or idea you've thought about or discussed? How do you think it relates to what we've talked about so far (evaluative thinking/evaluation culture)?

Conclusion

33. Those are all the questions I have. Is there anything else you'd like me to know about the EC in your workgroup, your work with professional evaluators, or the context of the organization or community that you think would be relevant?
34. Thank you for dedicating your time and experience to this study. I will be working over the few months synthesizing the information from the surveys, document review, and interviews. As this work comes together, I will also be collating the summary recommendations I mentioned earlier. I hope to be able to provide these to phase two participants by [INSERT DATE]. If these re not ready at that time, I will be sure to drop you a note to let you know when to expect it.
35. If you have any further questions or input, my contact information is on the screen.
36. If you have any concerns about this study, you can contact my advisor or the USF IRB. This contact information is also on the screen.

Appendix I: Response Summary

Included in request for phase two participation to provide value for survey participation and to help prepare for the case-study interview. Formatting preserved.

Participant

Name:

Institution:

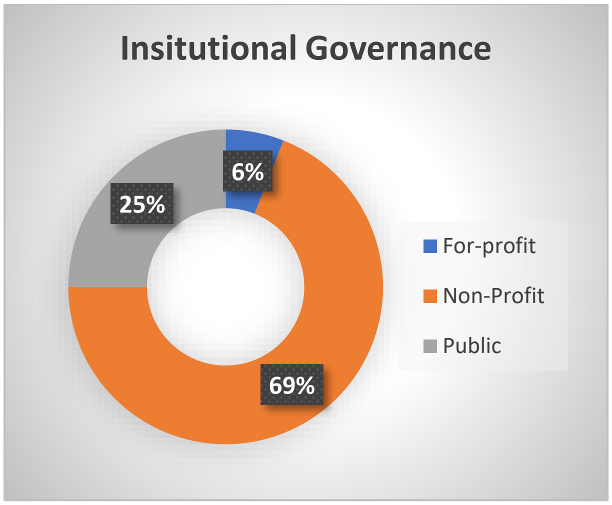
Date of Completion:

Institutional Demographics

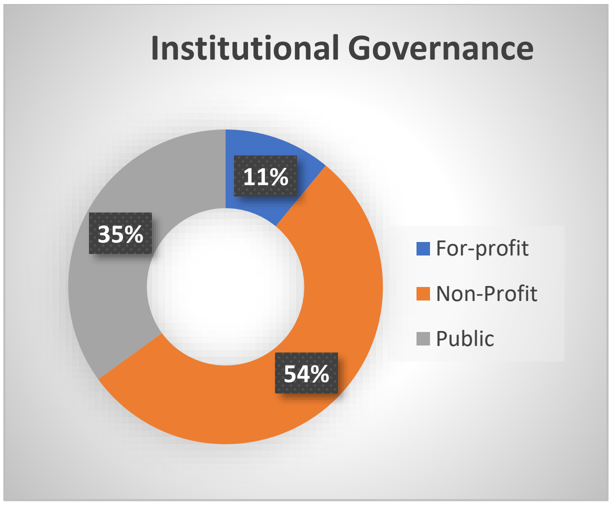
There were 100 institutional responses in the survey data set. What follows is a summary of those demographics compared to the broader AZA community. Study parameters specified AZA accreditation, combined multi-site institutions with joint management structures, and excluded international institutions, ultra-large corporate institutions, and institutions with conflicts. The later institutions are included in the AZA comparison because the statistics are drawn from AZA annuals reports that only include anonymized organizational information.

Governance

Response:



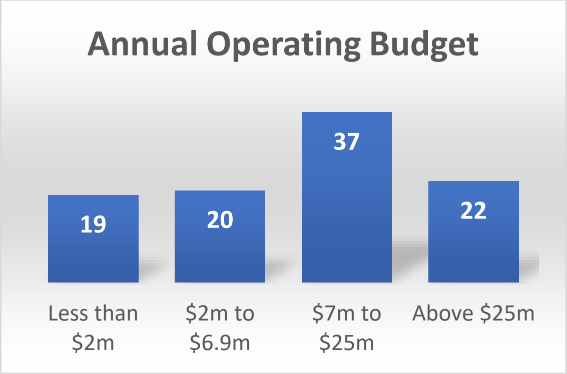
Sample (n = 100)

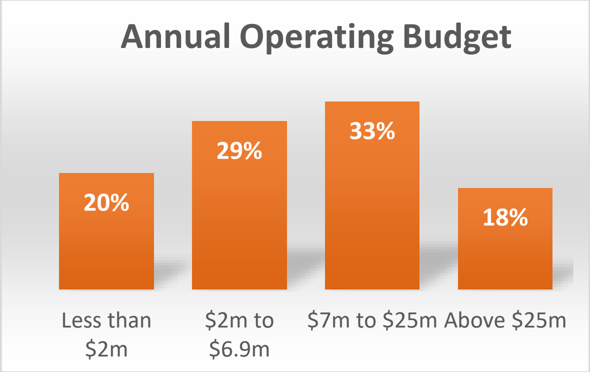


AZA (N = 240)

Annual Operating Budget

Response:



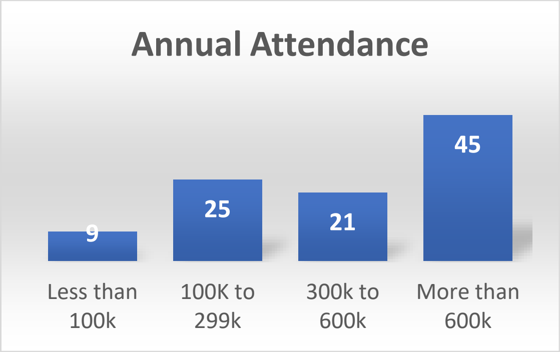


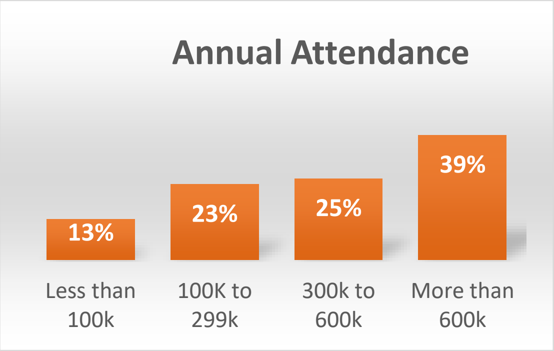
Sample (n = 100)

AZA (N = 240)

Annual Attendance

Response:



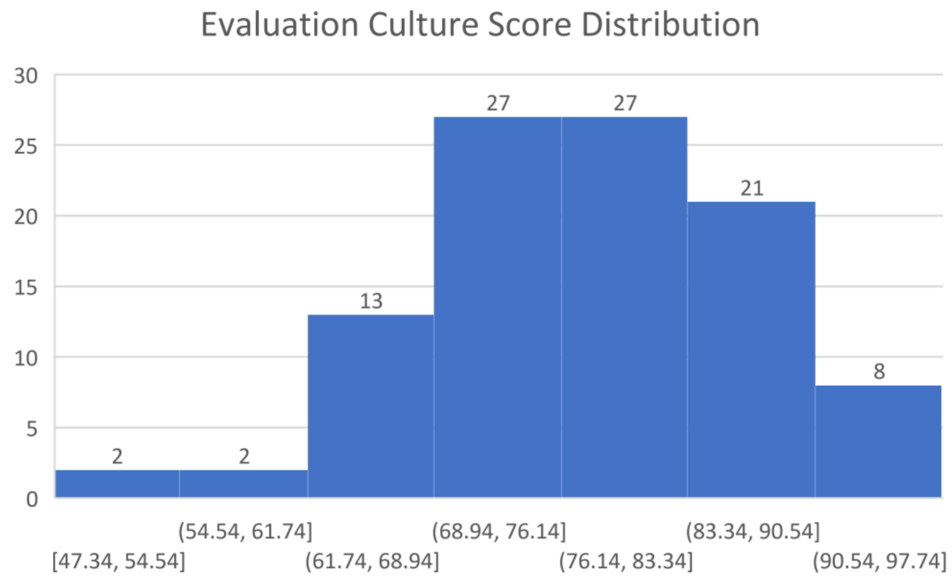


Score Card

You answered 50 items related to organizational learning and evaluation. These items were modified from the Readiness for Organizational Learning and Evaluation (ROLE) instrument (Preskill & Torres, 2000) and closely mirror the two constructs of interest in this study: evaluative thinking and evaluation culture. The items were divided into six subscales of variable sizes: Culture, Leadership, Systems & Structures, Communications, Teams, and Evaluation. A cumulative average was calculated for each scale and for the total score. The total score was calculated as an average of averages to prevent larger subscales from overly influencing the final number. Your scores on the six subscales and your total score are provided below along side the average scores from the 100-institution study sample. Each item was score from 1-100 so the mean scores will reflect a similar range.

Your Score	Sample Means
Culture (20 items)	
	79.96
Leadership (8 items)	
	80.64
Systems & Structures (7 items)	
	74.92
Communication (2 items)	
	60.79
Teams (5 items)	
	74.74
Evaluation (8 items)	
	77.68
Evaluation Culture (average of 6 subscale means)	
	74.78

Score Distribution

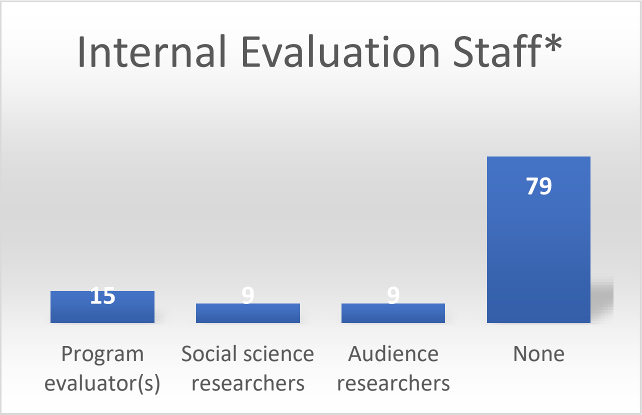


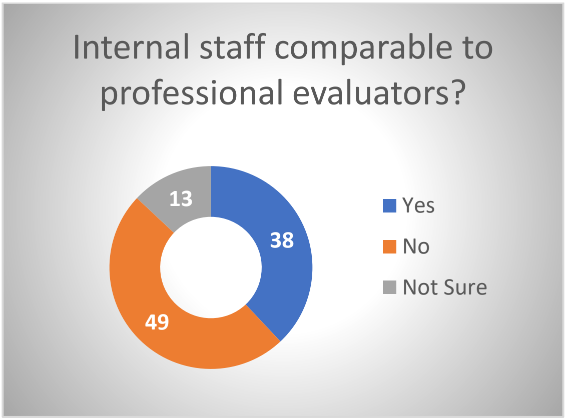
Work with Professional Evaluators

The hypothesis for the study was that institutions that more commonly work with professional evaluators would have a stronger evaluation culture (higher Evaluation Culture total score above).

Internal Evaluation Staff

Response(s):

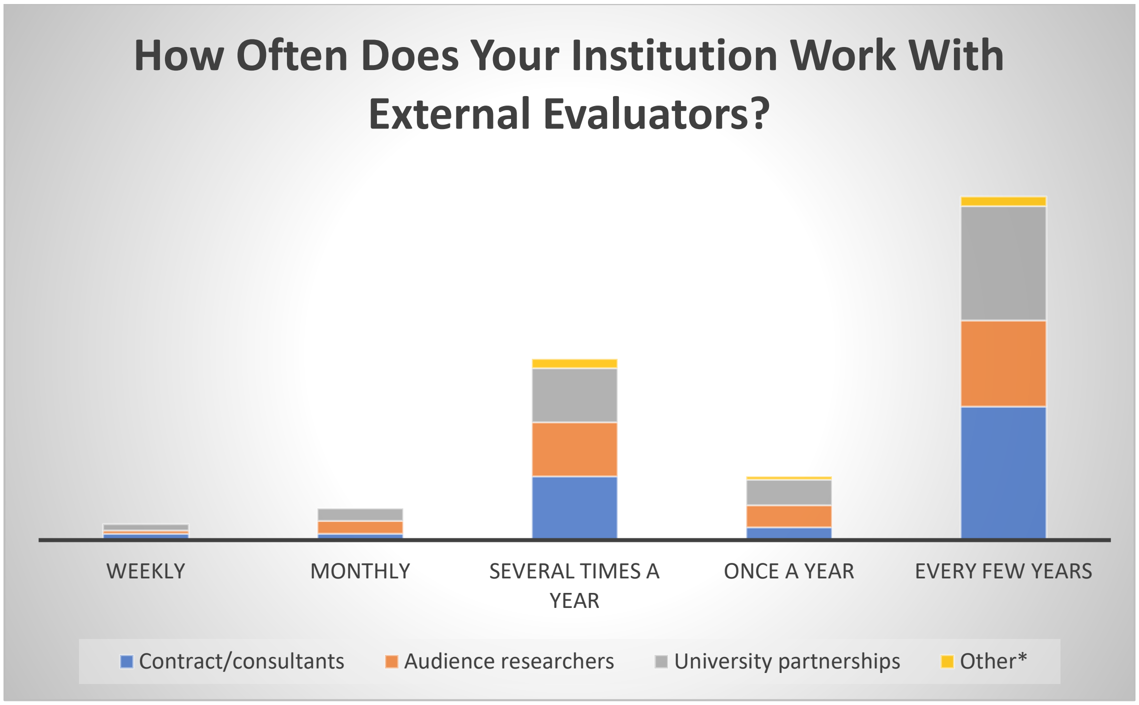




*Some institutions may have indicated multiple choices

Work with External Evaluators

Response:



Interview Questions

Here are the questions and topics we will discuss in our interview. Our discussion will be conversational so exact wording may vary. Throughout our interview, you are welcome to ask your own questions and clarifications about any of the study information or results. I would like to record our interview for note-taking purposes.

The overall goal of the study is to begin to understand the relationship between the way we work with professional evaluators and the evaluation culture of an organization (specifically the programming departments of accredited zoos and aquariums).

1. I will share with you the study definition of a 'professional evaluator' and we will discuss your work with staff who meet this definition. This could include staff who work in your department, elsewhere in your organization, or external staff you work with in some capacity (consultants, researchers, students, etc.). We will also discuss whether you or any of your staff have professional training or experience in evaluation.
2. I will ask you your thoughts about what it means to have an 'evaluation culture' in a workgroup or organization. I will share the study definition of evaluation culture and we will discuss how various dimensions covered in the survey may or may not influence the evaluation culture in your workgroup/department. We will also discuss whether/how you think your staff would view the evaluation culture of your department (through the lens of the survey) differently. We will also discuss if/how your work with professional evaluators may have influenced your department's evaluation culture.
3. Similarly, I will ask your thoughts about 'evaluative thinking,' share the study definition, and discuss. We will talk about key elements in the construct and your view on how they show up in your workgroup/department staff's work. Again, we will discuss how your work with professional evaluators may have influenced your staff's ability to think evaluatively.
4. Finally, we will discuss an interesting construct that emerged from the survey responses: psychological safety. Edmonds (1999) defines psychological safety as, "a shared belief held by members of a team that the team is safe for interpersonal risk-taking." I'm interested to learn from you your view on the role of psychological safety within your workgroup and how you view its influence on team learning and performance.

Survey Questions

Here are the questions you answered in the survey. They were randomized by the survey software.

Organizational Culture

1 Department employees respect each other's perspectives.

2 Department employees ask each other for information about work issues and activities.

3 Department employees continuously look for ways to improve processes, products and/or services.

4 Department employees are provided opportunities to reflect on their work.

5 Department employees often talk about the pressing work issues we're facing.

6 When trying to solve problems, department employees use a process of working through the problem before identifying solutions.

7 Department employees operate from a spirit of cooperation, rather than competition.

8 Department employees tend to work collaboratively with each other.

9 Mistakes made by department employees are viewed as opportunities for learning.

10 Department employees continuously ask themselves how they're doing, what they can do better, and what is working.

11 Department employees are confident that mistakes or failures will not affect them negatively.

12 Managers and supervisors in the department view individuals' capacity to learn as among the organization's greatest resources.

13 Department employees use data/information to inform their decision-making.

14 Asking questions and raising issues about work with department leaders is encouraged.

15 Department employees are not afraid to share their opinions even if those opinions are different from the majority.

16 Department employees feel safe explaining to others why they think or feel the way they do about an issue.

17 Department employees are encouraged to take the lead in initiating change or in trying to do something different.

18 Managers and supervisors in the department make decisions after considering the input of those affected.

19 In meetings, department employees are encouraged to discuss the values and beliefs that underlie their opinions.

20 Department employees are encouraged to offer dissenting opinions and alternative viewpoints.

Leadership

21 Managers and supervisors in the department take on the role of coaching, mentoring and facilitating employees' learning.

22 Managers and supervisors in the department help employees understand the value of experimentation and the learning that can result from such endeavors.

23 Managers and supervisors in the department are open to negative feedback from employees.

24 Managers and supervisors in the department model the importance of learning through their own efforts to learn.

25 Managers and supervisors in the department believe that success depends upon learning from daily practices.

26 Managers and supervisors in the department support the sharing of knowledge and skills among employees.

27 Managers and supervisors in the department provide the necessary time and support for systemic, long-term change.

28 Managers and supervisors in the department use data/information to inform their decision-making.

Systems and Structure

29 There is little bureaucratic red tape when trying to do something new or different in the department.

30 There are few boundaries between department units or working groups that keep employees from working together.

31 Department employees are recognized or rewarded for learning new knowledge and skills.

32 Department employees are recognized or rewarded for helping solve organizational problems.

33 The current reward or appraisal system in the department recognizes, in some way, team learning and performance.

34 Employees are recognized or rewarded for helping each other learn.

35 Department employees are recognized or rewarded for experimenting with new ideas.

Communication

36 Information is gathered from clients, customers, suppliers or other stakeholders during department activities to gauge how well we’re doing.

37 There are adequate records of past change efforts and what happened as a result.

Teams

38 Our department currently operates via (or is transitioning towards) a team-based structure.

39 Department employees are provided adequate training on how to work as a team member.

40 Team meetings in the department address both team processes and work content.

41 Team meetings in the department strive to include everyone’s opinion.

42 Teams and work groups in the department are encouraged to learn from each other and to share their learning with others.

Evaluation

43 The integration of evaluation activities into our department's work has enhanced (or would enhance) the quality of decision-making.

44 Managers and supervisors in the department like (or would like) us to evaluate our efforts.

45 Evaluation helps (or would help) the department provide better programs, processes, products and/or services.

46 There would be support among department employees if we tried to do more (or any) evaluation work.

47 Doing (more) evaluation would make it easier to convince department and organizational leadership of needed changes.

48 There are evaluation processes in place that enable department employees to review how well changes we make are working.

49 When the department engages in evaluation activities, the goal is to improve programs.

50 Data are routinely collected during department activities to inform evaluation efforts.

Appendix J: Statistical Supplement

Table J1

Correlation Matrix for Dimensional Means from Modified ROLE Instrument (N=100)

Dimension		Org Cul	Lead	Systems	Comm	Teams	Evaluation
Organizational culture	Pearson	1.00					
	Sig.	-					
Leadership	Pearson	.76	1.00				
	Sig.	.00	-				
Systems & structures	Pearson	.70	.63	1.00			
	Sig.	.00	.00	-			
Communication	Pearson	.48	.45	.39	1.00		
	Sig.	.00	.00	.00	-		
Teams	Pearson	.67	.63	.57	.43	1.00	
	Sig	.00	.00	.00	.00	-	
Evaluation	Pearson	.57	.56	.54	.48	.50	1.00
	Sig.	.00	.00	.00	.00	.00	-

Note. All correlations significant at the 0.01 level (2-tailed). Org cul = organizational culture. Lead = Leadership. Systems = systems & structures. Comm = communication.

Table J2

Pattern Matrix Exploratory Factor Analysis

	Factor			
Item	1	2	3	4
CUL1	-.097	.138	.693*	-.080
CUL2	.329	-.245	.291	.303
CUL3	.439*	-.115	.244	.181
CUL4	.646*	.217	-.093	-.042
CUL5	.104	-.087	.259	.256
CUL6	.516*	-.192	.160	.283
CUL7	-.061	.115	.721*	-.138
CUL8	.269	-.133	.764*	-.135

Table J3

Structure Matrix Exploratory Factor Analysis

	Factor			
Item	1	2	3	4
CUL1	.252	.429	.685	.248
CUL2	.517	.212	.452	.496
CUL3	.600	.309	.472	.475
CUL4	.681	.448	.311	.372
CUL5	.326	.225	.378	.388
CUL6	.656	.271	.429	.543
CUL7	.259	.410	.692	.211
CUL8	.495	.339	.757	.293

Table J2 (continued)

	Factor			
Item	1	2	3	4
CUL9	.029	.565*	.142	.005
CUL10	.613*	.015	.106	-.051
CUL11	-.134	.404*	.374*	-.102
CUL12	.362*	.335	-.035	.040
CUL13	.811*	-.127	-.002	-.109
CUL14	-.053	.721*	-.043	.042
CUL15	-.095	.212	.614*	-.012
CUL16	.034	-.094	.832*	-.056
CUL17	-.018	.373*	.092	.311
CUL18	.064	.404*	.171	.013
CUL19	-.032	.184	.114	.329
CUL20	-.163	.612*	.065	.265
L1	.422*	.225	.258	-.237
L2	.528*	.261	.056	-.012
L3	-.016	.834*	.108	-.258
L4	.366*	.505*	.001	-.063
L5	-.029	.147	.291	.084
L6	.096	.322	.290	.010
L7	.417*	.269	.011	.162
L8	.785*	-.022	.027	-.177
SS1	-.057	.210	.307	.059
SS2	.136	.169	.179	.171
SS3	.273	.111	.045	.416*
SS4	.340	-.044	.079	.214
SS5	.430*	.070	.157	.060
SS6	-.037	.774*	-.016	-.050
SS7	.265	.030	.131	.362*
COM1	.905*	-.086	-.214	-.124
COM2	.602*	-.234	.105	.047
T1	.368*	.106	-.080	-.027
T2	.564*	-.029	.157	.009
T3	.522*	.324	-.047	-.068
T4	-.243	.281	.567*	.193
T5	.048	.061	.175	.604*
E1	.131	.112	-.439	.805*
E2	.106	.361*	-.029	.206
E3	-.106	-.060	-.078	.802*
E4	-.099	.254	-.130	.440
E5	-.203	-.142	.039	.639*
E6	.842*	.011	-.257	.016

Note. Extraction method: Principal Axis Factoring, Rotation method: Promax with Kaiser normalization. Dimensions abbreviations used in item labels: organizational culture (CUL), leadership (L), systems and structures (SS), communication (COM), teams (T), evaluation (E).

* Factor loadings >.36

Table J3 (continued)

	Factor			
Item	1	2	3	4
CUL9	.362	.658	.463	.357
CUL10	.643	.334	.382	.338
CUL11	.176	.494	.483	.189
CUL12	.523	.504	.335	.382
CUL13	.691	.197	.264	.270
CUL14	.285	.693	.340	.341
CUL15	.288	.494	.678	.317
CUL16	.354	.345	.773	.295
CUL17	.368	.564	.426	.523
CUL18	.340	.533	.426	.320
CUL19	.286	.390	.348	.452
CUL20	.297	.699	.438	.500
L1	.520	.447	.473	.218
L2	.669	.531	.441	.426
L3	.283	.760	.434	.184
L4	.568	.646	.419	.380
L5	.223	.331	.395	.271
L6	.389	.529	.514	.349
L7	.635	.547	.427	.523
L8	.691	.272	.306	.251
SS1	.219	.378	.421	.269
SS2	.393	.411	.412	.408
SS3	.572	.463	.423	.638
SS4	.473	.260	.313	.413
SS5	.570	.384	.426	.399
SS6	.289	.724	.362	.296
SS7	.538	.399	.436	.580
COM1	.695	.159	.111	.229
COM2	.569	.126	.285	.309
T1	.365	.221	.139	.188
T2	.629	.322	.412	.372
T3	.613	.509	.344	.350
T4	.261	.567	.691	.453
T5	.487	.469	.505	.739
E1	.413	.324	.048	.731
E2	.372	.493	.309	.424
E3	.266	.236	.203	.681
E4	.197	.350	.160	.450
E5	.097	.093	.156	.478
E6	.735	.272	.155	.363

Note. Extraction method: Principal Axis Factoring, Rotation method: Promax with Kaiser normalization.Dimensions abbreviations used in item labels: organizational culture (CUL), leadership (L), systems and structures (SS), communication (COM), teams (T), evaluation (E).

Table J4

Factor Correlation Matrix for Exploratory Factor Analysis

Factor	1	2	3	4
1	1.00			
2	.47	1.00		
3	.47	.54	1.00	
4	.54	.48	.45	1.00

Note. Extraction method: Principal Axis Factoring, Rotation method: Promax with Kaiser normalization

Table J5

ANOVA Results for Study Variables on New Overall Evaluation Culture Score[a]

	F	Between Groups *df*	Within Groups *df*	*p*
Institutional governance[b]	0.07	2	97	.93
Operational Budget[c]	0.28	3	94	.84
Annual Attendance[c]	1.90	3	96	.14
Internal Evaluators[d]	0.43	1	98	.52
External Evaluators[d]	1.20	1	98	.28
Trained Internal (non-evaluator) staff[d]	1.78	1	98	.19
Combinations[e]	1.13	6	93	.35

Note. [a]An average of the four new dimensional means (evaluative thinking, evaluation/growth, leadership-related psychological safety, team-related psychological safety) associated with the four factors identified in an exploratory factor analysis. [b]For-profit, non-profit, public. [c]Small, medium, medium-large, large. [d]Presence/ absence. [e]Seven total combinations of internal evaluators, external evaluators, and trained staff (including none).

Table J6

MANOVA Results for Study Variables on New Dimensions[a]

		Between	Within			
	F	Groups *df*	Groups *df*	*p*	Λ	η^2
Institutional governance[b]	1.74	10	186	.07	.84	.09
Operational Budget[c]	1.12	15	249	.34	.84	.06
Annual Attendance[c]	1.65	15	254	.06	.77	.08
Internal Evaluators[d]	2.56	5	94	.03	.88	.12
External Evaluators[d]	4.13	5	94	.00	.82	.18
Trained Internal (non-evaluator) staff[d]	1.05	5	94	.39	.95	.05
Combinations[e]	3.23	25	332	.00	.45	.15

Note. [a]An average of the four new dimensional means (evaluative thinking, evaluation/growth, leadership-related psychological safety, team-related psychological safety) associated with the four factors identified in an exploratory factor analysis. [b]For-profit, non-profit, public. [c]Small, medium, medium-large, large. [d]Presence/ absence. [e]Seven total combinations of internal evaluators, external evaluators, and trained staff (including none).

Table J7

Results of Multiple Regression on New Overall Evaluation Culture

Variable	*B*	*SE B*	β	*p*
For-profit[a]	-8.20	5.39	-.17	.13
Public	-0.55	2.42	-.03	.82
Budget	0.45	1.14	.05	.70
Evaluator Conditions[b]				
No-No-No[c]	7.86	4.50	.20	.08
Yes-No-No	16.43	9.88	.17	.10
Yes-Yes-No	2.19	4.02	.06	.59
Yes-Yes-Yes	0.19	3.50	.01	.96
No-Yes-Yes	4.64	2.50	.21	.07
No-No-Yes	2.18	5.83	.04	.71
$R^2 = .10$				

Note. [a]Reference group for Governance is Non-profit. [b]Evaluator conditions indicate presence/absence of internal evaluations-external evaluators-trained (non-evaluator) staff. [c]Reference group for Evaluator Conditions is No-Yes-No.

About the Author

Jim Wharton is the Director of Conservation Engagement and Learning at the Seattle Aquarium. He has been in the field of conservation education and engagement since 1997 with interests in developing empathy for ocean animals and fostering an ocean ethic in our community. His research interests are varied and include the role of empathy in developing conservation outcomes, how public perceptions of sharks contribute to their conservation, and how evaluators contribute to the evaluation culture of an organization. Before Seattle, Jim worked at Mote Marine Laboratory and Aquarium, the Smithsonian Marine Station, and the Oregon Coast Aquarium. Jim has a B.S. in biology from the University of Michigan and a M.S. in Marine Resource Management from Oregon State University.