Journal of MultiDisciplinary Evaluation
Volume 13, Issue 29, 2017

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Communicating About Evaluation: A Conceptual Model and Case Example[1]

Bianca Montrosse-Moorhead
University of Connecticut

Aarti P. Bellara
University of Connecticut

Anthony J. Gambino
University of Connecticut

Background: Despite consensus within the evaluation community about what is distinctive about evaluation, confusion among stakeholders and other professions abounds. The evaluation literature describes how those in the social sciences continue to view evaluation as applied social science and part of what they already know how to do, with the implication that no additional training beyond the traditional social sciences is needed. Given the lack of broader understanding of the specialized role of evaluation, the field struggles with how best to communicate about evaluation to stakeholders and other professions.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Not applicable

Purpose: This paper addresses the need to clearly communicate what is distinctive about evaluation to stakeholders and other professions by offering a conceptual tool that can be used in dialogue with others. Specifically, we adapt a personnel evaluation framework to map out what is distinctive about what evaluators know and do. We then compare this map with the knowledge and skill needed in a related profession (i.e., assessment) in order to reveal how the professions differ.

Findings: We argue that using a conceptual tool such as the one presented in this paper with comparative case examples would clarify for outsiders the distinct work of evaluators. Additionally, we explain how this conceptual tool is flexible and could be extended by evaluation practitioners in a myriad of ways.

Keywords: *evaluation knowledge; evaluation skill; profession; professionalization.*

[1] An earlier draft of this paper was presented at the annual American Evaluation Association conference, Atlanta, GA, October 2016.This article has benefited greatly from feedback and encouragement from Drs. Suzanne Wilson and Chris L. S. Coryn, and the anonymous reviewers. We also wish to thank Brittney Hernandez with her assistance in preparing the reference list for this paper. Any remaining errors or omissions are our own.

Correspondence concerning this article should be addressed to Bianca Montrosse-Moorhead, Ph.D., Assistant Professor of Measurement, Evaluation, and Assessment, Department of Educational Psychology, University of Connecticut, 249 Glenbrook Rd., U-3064, Storrs, CT 06269. Email: bianca@uconn.edu

Introduction

The demand for high quality, actionable, and professional evaluation has grown exponentially. Indicators lending support for this claim include, for example, that evaluation is being conducted on multiple types of evaluands across multiple contexts (Donaldson, 2013). There is now general consensus that part of what makes evaluation distinct from other forms of disciplined inquiry is the specialized expert knowledge we bring to bear in conducting evaluations aimed at addressing social problems. On December 19, 2014, UN resolution 69/237 was approved, a significant milestone because it is the first UN resolution to call for evaluator and evaluation capacity at the country level. Grassroots movements, like EvalPartners, have documented more than 100 Voluntary Organizations of Professional Evaluators (VOPEs)[2], which suggests that formal spaces for meeting other evaluators, building social capital, sharing information, and learning are available worldwide. And, increased opportunities for acquiring evaluation knowledge, skills, and dispositions exist, for example, through graduate programs, professional development workshops, certificate programs, webinars, and toolkits (LaVelle, & Donaldson, 2015; Schwandt, 2015).

Concurrently and related, the conversation on Evaluation as a profession and a discipline has once again picked up momentum, especially in regards to policing mechanisms such as accreditation, certification, competencies, credentialing, licensing, badges, and peer review (Altschuld & Engle, 2015; Davies, Randall, & King, & Stevahn, 2015; Oksanen, 2016; West, 2015). Inherent in these debates are two fundamental questions: what is distinctive about what evaluation practitioners know and do, and how can that information best be used in dialogue with stakeholders and other professionals to promote shared understanding? This paper addresses both of these questions.

In this paper, it is argued that despite consensus within evaluation about the unique work of evaluators, there is still a legitimate and important need to clearly communicate what is distinctive about evaluation in ways that stakeholders and other professionals understand. This need is tied to larger conversations about how to maintain Evaluation's professional status, and how to reinforce our disciplinary boundaries. To address this need, we present a communication tool intended to be used with outsiders. To promote use of this tool, a case example is presented in which the knowledge and skills needed by an evaluator are compared to those needed by an assessment expert. Implications of the use of this conceptual tool are also explored.

Articulating What Makes Evaluation Distinct is Important for the Discipline and Profession

Several reasons exist as to why being able to clearly explain what is distinct about Evaluation is important to pursue. One, being able to do so is tied to our status as a discipline and profession (Fournier, 1995, 2005; House, 1977; 1980; Scriven, 1991, 1993, 2007). And two, there is legitimate concern within the broader field that professionals from other disciplines are claiming the title of "evaluator" and conducting low-quality evaluations, which has repercussions for how outsiders view Evaluation. In what follows, these two reasons are discussed in greater depth, with several key terms defined.

Evaluation's Status as a Profession

In what follows, we first define how the terms "professionalize" and "profession" are used in this paper to relate current debates about the professionalization of evaluation back to larger sociological discussions about professions generally. Next, we bear down on whether evaluation meets the criteria for a profession. In this discussion, we offer a novel perspective on one of these criteria – whether evaluation polices itself – as views on how evaluation meets this criterion has been the subject of much debate.

According to the Oxford English Dictionary, the term *professionalize* is action-oriented and means to "give (an occupation, activity, or group) professional qualities, typically by increasing training or raising required qualifications" (2017). Sociologists have long studied what activities groups of people and occupations engage in to increase their professional qualities, including those that are used to develop and define professions (Macdonald, 1995; Millerson, 1964; Freidson, 2001). *Professions* are paid occupations, which require extended training and formal qualifications (Oxford English Dictionary, 2017). Sociologists now generally agree that five common characteristics are observed across professions:

[2] http://www.ioce.net/vope-directory

they have programs where extended training occurs; they have unique specialized knowledge; they have codified ethical principles; they exercise autonomy because of their unique specialized knowledge and formal qualifications; and they are self-policing (Picciotto, 2011; Freidson, 2001).

Long-standing debates in evaluation on whether it is a profession exist. Some luminaries argue that evaluation is a profession (Patton, 1990; House, 1993), some that it is a developing profession (Lincoln, 1985; Picciotto, 2011), and others that it is not a profession (Rossi, Lipsey, & Freeman, 2004).

Our perspective is that evaluation is a profession because the same five characteristics are observed. The growth of university-based evaluation training programs is well documented (LaVelle, 2014), as is the proliferation of other training opportunities to expand upon formal training received in universities, for example, workshops and short courses (LaVelle, & Donaldson, 2010). These university programs and other training opportunities routinely aim to deepen evaluators' understanding and application of the specialized knowledge evaluators bring to bear in their practice. This includes knowledge developed in evaluation, such as the practical application of evaluation theories and approaches, evaluability assessment, and evaluation synthesis methodology (Dewey, Montrosse, Schröter, Sullins, & Mattox, 2008; LaVelle, 2014). It also includes specialized knowledge developed outside of evaluation, for example, research methods, but which still requires extended training. Furthermore, evaluators who have formal training in evaluation and practice in the United States are guided by the *Guiding Principles for Evaluators* (American Evaluation Association, 2004), an ethical code of conduct, while those that practice internationally are guided by contextually- and culturally-bound ethical principles developed by other international evaluation associations (e.g., ANZEA, & SuPERU, 2015; Rodríguez Bilella, Martinic Valencia, Soberón Alvarez, Klier, Guzmán Hernández, & Tapella, 2016; Patel, 2013). Because of the formal training, the specialized knowledge evaluators possess, and the ethical principles that guide our work, generally speaking, we are afforded autonomy to do our work. That is, a body outside of our professional walls does not regulate evaluation practice. Although, it is important to note that like all professions, this autonomy is always under threat (Freidson, 2001). Rather, we are self-policing. It is this last point, specifically with respect to *how* evaluators ought to self-police, which has received much attention, and which we discuss next.

Self-Policing Mechanisms in Evaluation

Multiple levels of self-policing have been proposed, which include accreditation, certification, credentialing, licensure, and badges (Altschuld, & Engle, 2015; Davies et al., 2015; Oksanen, 2016). *Accreditation* is a formal process, which is based on an external review of a preparation or educational program to see that it meets profession or discipline standards. *Certification* is usually determined by performance on a test that purports to measure a person's competence on specific skills and knowledge. *Badges* utilize a micro-certification strategy with a top-down design process that breaks up the larger certification problem into flexible, manageable units (Davies, Randall, & West, 2015). *Credentialing* is a system that results in a credential or designation of such status, and is intended to convey that a professional has the requisite knowledge, skills, and practical experiences needed. *Licensure* is traditionally conferred through a governing body or under the aegis of the government, and aims to convey that a professional has a right to practice. *Voluntary evaluator peer review* is a recent phenomenon, spearheaded by the European Evaluation Society (EES), and intends to promote evaluator capabilities and accountability[3].

How evaluators self-police is varied across the globe. For example, the American Evaluation Association has established a task force to develop a set of evaluator competencies, with a formal board vote to adopt the evaluator competencies expected in late 2017 (J. King, task force chair, personal communication). Europe has taken a different approach. In September 2015, the European Evaluation Society formally launched the Voluntary Evaluator Peer Review (VEPR) program. This program aims to improve evaluators' capabilities and professional accountability. Intentionally, the EES VEPR program was developed by evaluators for evaluators in recognition that only those with the specialized evaluation knowledge, skills and dispositions are in a position to value evaluator competency. Canada, Japan, and Thailand have established formal evaluator credentialing systems, which are overseen by evaluation associations in those countries (King, & Stevahn, 2015). Moreover, a recent study of the Canadian Credentialed Evaluator (CE) program found that

[3] http://www.europeanevaluation.org/events/ees-conferences-and-events/conferences/evalyear-2015x/vepr-project

members' attitudes and perceptions were generally positive, and CEs also attributed improvements in their practice and an increase in the sense of belonging to the evaluation profession as a result of the credential (Gauthier, Kishchuk, Borys, & Roy, 2015).

Two perspectives on how to read the evidence in regards to how evaluators self-police exist. Some scholars argue it is a lack of a self-policing mechanism on which evaluation fails to meet the criteria for a profession (Picciotto, 2011). This was based on the fact that, at the time, only the Canadian CE program existed. But, as we have described, much has changed in a short time, and more self-policing mechanisms are currently in place. Furthermore, evaluators living outside of Canada, but wishing to conduct evaluations in that country, have earned a Canadian CE credential (Gauthier, et al., 2015). For these reasons, we put forth an alternative perspective on the evidence in regards to how evaluators self-police. That is, we believe evaluators are self-policing, which also means that Evaluation meets all of the criteria for a profession.

Evaluation's Status as a Discipline

Having made the case that Evaluation is a profession, in what follows the case is made that evaluation is also a discipline. First, key ideas promulgated by Toulmin are reviewed, including criteria for identifying disciplines, and an example of how the social sciences meet these criteria. Then, the criteria are applied to Evaluation to make the case for its disciplinary status claim.

Whereas professions are comprised of a complementary set of the five characteristics discussed above with a focus on organizations, responsibilities, and practitioners, a discipline is comprised of methods for dealing with both theoretical and practical problems (House, 1993). Toulmin (1972) defines a *profession* as:

> A collective human enterprise takes the form of a rationally developed "discipline," in those cases where men's [and women's] share commitment to a sufficiently agreed set of ideals leads to the development of an isolatable and self-defining repertory of procedures; and where those procedures are open to further modification, so as to deal with problems arising from the incomplete fulfillment of those disciplinary ideals. (p. 359)

Toulmin also provides a framework for identifying disciplines. Briefly, these include having common goals, criteria and standards of agreement, professional forums, and a disciplinary structure.

The social sciences provide one example of how these criteria operate in the real-world. The social sciences hold explanation as their common goal. Methodological criteria and standards ideals exist and are used to judge the legitimacy and validity of published social science studies. Professional journals and organizations serve a central role in the exchanging of competing ideas. And, there is a clear disciplinary structure. For example, anthropology, economics, psychology, and sociology – all social sciences – study different aspects of the human experience in an attempt to explain it. As such, it is possible for individuals to become specialists in one of these areas through formal university training. Furthermore, one typically becomes a specialist in a particular branch; for example, in psychology, one can specialize in cognitive psychology, developmental psychology, social psychology, etc.

How does evaluation measure up against these criteria? The goal of evaluation is valuing, and clear, formal (Scriven, 1991), and working logics (Fournier, 1995, 2005) have been established and agreed upon. These formal and working logics are further codified in meta-evaluation literature (Cook & Gruder, 1978; Nilsson & Hogben, 1983; Scriven, 1975, 2009; Stufflebeam, 1974/2011, 1978, 2001), with meta-evaluative criteria and standards used to judge the legitimacy and validity of published evaluations. Several evaluation journals are well-established, such as this one, and others (e.g., *American Journal of Evaluation*, *New Directions for Evaluation*, *Evaluation and Program Planning*). The exchange and debate of ideas is further promoted through annual conferences of major evaluation associations (e.g., the American Evaluation Association), and the proliferation of evaluation associations internationally is well documented (Catsambas, & Bauer, 2015). Historically, in evaluation, this exchange has been marked by defining problems juxtaposed against the ideals of the time and solving those problems. This whole enterprise is context-bound, meaning that ideals, problems, and solutions are always in flux and bear the marker of their place and time. This is also consistent with patterns observed in other disciplines (House, 1993). And, our disciplinary structure has been established. Evaluation can be further sub-divided into different branches, for example, program evaluation, product evaluation, personnel evaluation, and so on (Scriven, 2015). All are united by the common goal of valuing. Further, while there is no department of evaluation, although Western Michigan

University's interdisciplinary evaluation program is close, it is possible for evaluators to receive formal training in universities in these different branches (LaVelle, & Donaldson, 2010, 2015). Clearly then, evaluation bears all of the markers to claim its disciplinary status.

Maintaining Evaluation's Designation as a Profession and Discipline

If Evaluation as a discipline cares about maintaining its status as a profession, attention must be paid to how definitional criteria are maintained, as neither the knowledge that we use in practice nor the contexts in which we work are static. Two of these criteria – the unique specialized knowledge evaluators possess and the degree of autonomy we enjoy because of that knowledge – are particularly important. This is because the unique specialized knowledge we bring to bear needs to be known and understood not only within our disciplinary and professional walls, but also by outsiders. After all, it is outsiders who create the demand for our particular brand of specialized knowledge, not the other way around. Further, evaluation professionals are only able to exercise autonomy, in part, because of how outsiders view our unique specialized knowledge.

For these reasons, the concerns expressed by evaluation scholars about who is meeting the demand for evaluation and how, especially in light of growing demand for evaluation, is legitimate. Evidence of this growing demand can be identified worldwide. For example, Porter and Goldman (2013) describe how government mandates have led to the creation of country-level evaluation systems in six African countries (Benin, Burundi, Ghana, Kenya, Senegal, South Africa and Uganda). Similar case examples are beginning to emerge from East Asia, Latin America, and from the Middle East and North African regions (EvalPartners, 2017; Segone, 2010). Further, UN resolution 69/237 is the first approved resolution to call for evaluator and evaluation capacity at the country level (Catsambas, & Bauer, 2015).

Several problems have surfaced in regards to increased demand. First, Schwandt (2015) has noted that, as a result of increased demand, "many who take on the job of conducting or managing an evaluation lack formal training or experience, resulting in evaluations that are poorly conceived, poorly executed, and poorly managed" (p. 128). Second, it is not completely clear if increased demand is actually for monitoring, a continuous improvement function aimed at taking stock of inputs and targets (OECD, 2002), or if increased demand is for evaluation. Some have noted the tendency for monitoring to masquerade as evaluation (Picciotto, 2010; Porter and Goldman, 2013). Third, it is not uncommon for the work of auditors, economists, organizational psychologists, and other applied professionals to be confused with what evaluation professionals can and are trained to do (Davidson, 2007; Jacob, 2008; Jacob, & Boisvert, 2010; Perrin, 2005). The concern here is that those who commission evaluations do not understand what is distinct about what evaluation professionals know and can do, which most immediately can lead to the concerns noted by Schwandt. And, evaluations that are poorly conceived, executed, and managed threaten the degree of autonomy experienced by evaluators, and thus, our professional status.

Communicating About What Makes Evaluation Distinct

Having made the case for why clarity about what makes evaluation distinct is important to pursue, two corresponding ideas are put forth in this section. One is that there is consensus within the evaluation community about what is distinctive about evaluation. The second is that that Evaluation still struggles with how best to communicate about evaluation to stakeholders and other professions.

After decades of debate about what evaluation is and what makes the work of practitioners distinct, a reasonable degree of consensus has been reached. It is now commonly understood how evaluation uses the methods of applied social science, and at the same time is distinct from it (Blome, 2009; Coffman, 2003/2004; Department of Health and Human Services, nd; Mathison, 2007). For example, there is a general and working logic of evaluation (Fournier, 1995, 2005; Scriven, 1991). Evaluation-specific methods, such as synthesis methods and criteria development methods, have been developed (Davidson, 2015; King, McKegg, Oakden, & Wehipeihana, 2013; Nunns, & Roorda, 2010). A substantial knowledge base on different evaluation approaches exists (Alkin, 2013; Shadish, 1999; Stufflebeam, & Coryn, 2014).

Yet, this common understanding within the field is not yet fully realized by evaluation stakeholders or other academic communities. Davidson (2007) provides an example of this confusion in describing how those in the social sciences continue to view evaluation as applied social science. She writes,

> Right across the social sciences and in many other disciplines where evaluation is relevant in more than just its intradisciplinary application, it seems that the vast majority of practitioners consider it to be part of their own toolkit already, albeit often under a different name. Most of these practitioners consider evaluators delusional when we suggest that evaluation is sufficiently distinct to call a profession, let alone an autonomous discipline. (p. 3-4)

Evaluation reports produced by government agencies are also instructive for understanding how stakeholders, including commissioners of evaluations, understand evaluation. In particular, the evaluation questions and methods used to answer those questions are particularly illuminating because in evaluation it is the stakeholders who pose questions for evaluators to answer. It is stakeholders' understanding of evaluation that guides the types of questions that are answered, and questions drive methods. Take for example, a recently released evaluation report on the DC Opportunity Scholarship Program (OSP; Dynarski, Rui, Webber, & Gutmann, 2017). This evaluation study was mandated by Congress and overseen by the National Center for Education Evaluation and Regional Assistance within the U.S. Department of Education. The purpose of this evaluation was to "address the impacts of being offered an OSP scholarship and the actual use of an OSP scholarship on (1) student achievement, (2) parent and student satisfaction, (3) parent- and student-reported school safety, and (4) parent involvement" (p. 3). Aligned with this purpose, the evaluation questions all centered on implementation and outcome questions. For example, one question posed asked about the effect of the OSP program on student language arts and mathematics achievement. What is striking about the questions posed is that none are evaluation questions because none are about the merit, worth, or significance of OSP implementation or outcomes. Rather, the questions are social science research questions, which make no attempt to *value* the OSP program. It is no surprise, then that the methods employed used social science methods, and no attempts were made to develop evaluative criteria, evaluative evidence, nor use those things to generate evaluative claims. To be clear, this is a well-designed and well-executed research study. An evaluation study it is not.

Given the lack of broader understanding among stakeholders and other professions of the specialized role of evaluation, it can be argued that the field struggles with how best to communicate about evaluation to stakeholders and other professions and that more work promoting evaluation is needed. One way to begin to systematically address this issue is through the development of a conceptual framework for mapping out what is distinct about evaluators. While much writing has been devoted to mapping out what is distinct about evaluation, few have used the evaluator as the unit of analysis for these discussions. We contend that shifting this focus to the evaluator is a good way to structure how we talk about what we do. This allows for outsiders to frame their ideas about what evaluators know and can do in their own personal contexts.

What might such a framework look like? It is this question we take up in the next section.

What Evaluators and Assessors Need to Know and be Able to Do

The previous section aimed to contextualize our paper in current literature on the professional status of evaluation, and make the claim that evaluation has met the criteria to be categorized as a profession. Within this discussion, we highlight the specialized knowledge evaluators bring to bear on their practice. In this section, we expand upon this criterion and begin to sketch out a framework for organizing this specialized knowledge. To do so, we adapt Scriven's personnel evaluation framework (1994, 2015).

A Framework for Categorizing the Specialized Knowledge of Evaluators

We ground this work in an adapted version of a personnel evaluation framework proposed by Scriven (1994, 2015), which focused on teacher evaluation. Scriven's central thesis was less catchy than the one put forth by George Bernard Shaw ("Those who can, do; those who can't, teach."), but arguably more correct: "Those who can do these hundred difficult things can teach well; those who can teach well can change the world in their lifetime; those who can't, will rarely do something as important."

To lay the foundation for his thesis, Scriven first argued that all practice-based professions could be outlined using three general anchoring questions: What are the generic duties of this practice-based profession?; How do the generic duties translate in a given context for a specific job?; and, how do evaluators in a given context for a specific job interpret the enactment of job-specific duties? In doing so, Scriven persuasively

drew connections between teaching, a new profession, and medicine and law, two of the original professions (Larson, 1977).

Few would disagree that evaluation is also a practice-based profession. The enactment of evaluation knowledge is precisely what we do when we evaluate. The same logic applies to assessment. The enactment of assessment knowledge is precisely what a practicing assessor does. Using the common thread of practice-based professions provides a rationale and entry point for adapting Scriven's original framework for our purposes (see Figure 1).

Within this framework, *generic duties* are defined as duties common to the particular practice-based profession. In the context of evaluation, this would be an acknowledgment of duties common to all evaluation jobs, and which, importantly, help to draw boundaries around the profession of evaluation and distinguish it from other professions. For example, knowledge of the landscape of prescriptive evaluation theories (or approaches) [4] would fall under generic duties (Alkin, 2013; Shadish, 1999). *Job-specific duties*, therefore, would encompass duties specific to a particular job. To illustrate, in the context of evaluation, an evaluator might choose to use a specific type of evaluation theory because it is appropriate for the context and purpose of that evaluation. However, in this example, use of a specific type of evaluation theory should not be confused with no need to know about or a lack of knowledge about other evaluation theories, as that is part of the generic duties. There is a third piece to this framework, *context-specific interpretation of each duty*, which should specify what will count as satisfactory performance within a particular context and which is used to guide evaluation efforts. Turning again to an evaluation example, this would specify the criteria to be used in meta-evaluative efforts of a particular evaluation in a particular context.

Furthermore, within each of these three dimensions, five domains are located. Scriven argued that these five domains represent the five areas *across* all practice-based professions in which it is generally agreed that a certain minimum level of competence is required to enact them, meaning that they designate when a professional is safe to practice. These five domains include: (i) subject matter knowledge, (ii) application skill, (iii) assessment skill, (iv) professionalism, and (v) "other" secondary duties.

Taking this portion of Scriven's framework, and adapting it to the duties of the evaluator, which is possible because, as already noted, evaluation is a practice-based profession, these same five distinct elements can be applied to evaluation. For example, *subject matter knowledge* in evaluation needs to be current, comprehensive, and accurate. A portion of this knowledge is context-independent knowledge, including knowledge that does not vary by context and is used in every evaluation or knowledge that has the potential to be used in every evaluation. For example, knowledge of synthesis methods would be an example of knowledge used in every evaluation, while knowledge of a broad range of evaluation theories of practice is an example of knowledge that has the potential to be used in a new evaluation setting. Moreover, both of these examples highlight elements that are unique to evaluation, and thus help distinguish it from other professions. Subject-matter knowledge is also comprised of emerging knowledge. Since knowledge is not static, this necessarily requires that professionals continually engage in some form of training (e.g. through workshops or coursework). An example of this type of knowledge is emerging work on development, use, and validation of evaluative rubrics as a type of synthesis method (Davidson, 2015; King, et al., 2013; Nunns, & Roorda, 2010).

Application skill is common across all practice-based professions, and includes competence in planning, project management, and communication. In the context of evaluation, this would include, for example, the ability to locate, respond to, and secure different types of requests leading to evaluation contracts. It's not uncommon for different commissioning organizations to request and require different types of documents that describe evaluation needs (e.g., terms of reference, scope of work, request for proposal, invitation to tender), and evaluators must be adept at being able to respond to these requests. Furthermore, once a contract is secured, the ability to develop and execute a project management plan is crucial. For example, an evaluator who is tasked with designing an outcome evaluation, and who chooses to use a Chenian theory-driven approach in doing so (Chen, 2015), will need to develop and execute a plan that allows

[4] Prescriptive evaluation approaches provide suggestions for carrying out a good evaluation (Alkin, 2004). Thus, current prescriptive evaluation approaches include guidance on how to conduct evaluations, based upon how evaluation scholars believe the evaluation *should be* conducted, not on how evaluations *are* conducted.

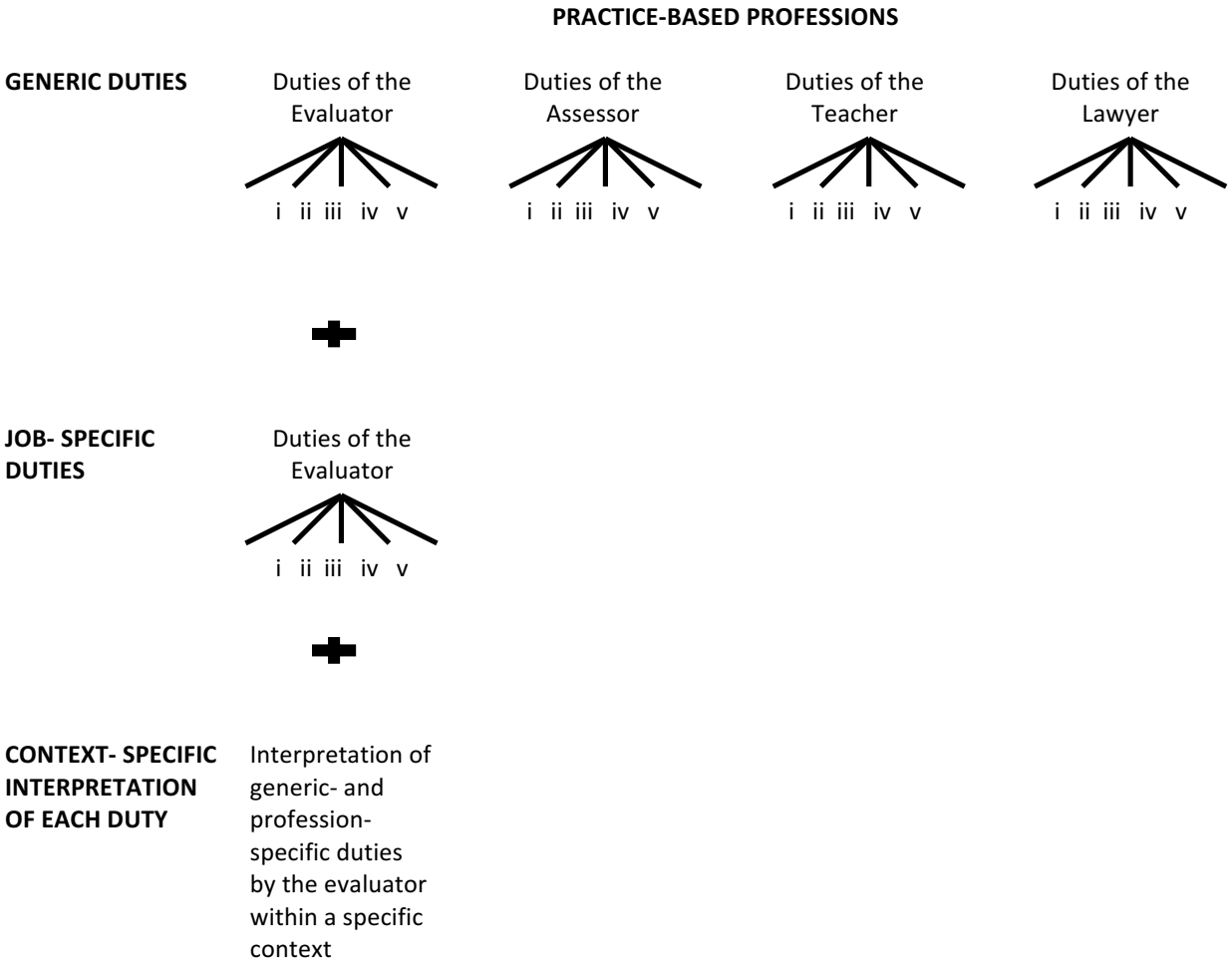

Figure 1. Mapping of the Duties of Evaluation Practitioners (Adapted from Scriven [1994, 2015]).

him or her to develop a program theory that identifies the underlying causal mechanisms, the implementation processes, and the outcomes, to develop a data collection plan that allows for the measurement of all of these program theory aspects, and to develop a data analysis plan using structural equation modeling to formally test hypothesized pathways between mechanisms, processes, and outcomes. Also important is the ability to create and implement a well specified communication plan, as communication is a factor that has been shown to influence the evaluation process (King, & Stevahn, 2013).

Knowledge about assessments, assessment construction, grading, ranking, and scoring practices as they relate to assessments, and the ability to record and report assessment information to a wide variety of stakeholders are also part of the minimum level of competence needed in practice-based professions; thus, providing a rationale that *assessment skills* are needed. Like Scriven (1994), we use the term assessment here in a broad sense, meaning we define it as "any systematic and objective process that leads to either *evaluative classification* … [or] *determination of the merit of [an evaluand]*" (p. 23, italics in original). In the context of evaluation, this would include things such as evaluators having the ability to be critical consumers of validity and reliability evidence for existing assessments, knowledge of factor analysis methods, knowledge of rules for scoring and interpreting scores from existing instruments, and the basics of survey design for circumstances in which assessments need to be created.

Professionalism touches on areas of minimum competence in practice-based professions surrounding professional ethics, interpersonal skills, service to the profession, knowledge about policy, and knowledge of context. In the context of

evaluation, this would include things like awareness and use of the *AEA Guiding Principles for Evaluators* (professional ethics). Also important are a minimum level of interpersonal competence. Service to the profession would include things like participating in formal leadership positions, such as serving as a Topical Interest Group leader within the American Evaluation Association (AEA), or in informal positions, such as serving as a mentor to a junior evaluation colleague. Knowledge of relevant policies, for example, policies governing acceptable evaluation designs at the U.S. Department of Education, is also paramount. Lastly, as articulated by Rog, Fitzpatrick, and Conner (2012), knowledge of context including the ways in which it can shape evaluation practice is important. For example, Fetterman has discussed a situation in which conditions signaled that an empowerment evaluation would not be feasible (Fitzpatrick, 2002), despite his association with this particular evaluation theory.

Lastly, there may be *other duties* required by practice-based professions that do not fit within the previous categories. In the context of evaluation, this would include things such as attending annual regional or national evaluation conferences; developing, knowing about, and reacting to policy changes (e.g., UN General Assembly Resolution 69/237); etc.

Having sketched out a general framework for mapping what evaluators need to know and be able to do, we now turn our attention to a specific case relevant for the current argument—the duties of the evaluation expert narrowed down to a context in which an evaluation seeks to answer an outcome question. This type of evaluation is typically designed to provide information regarding how valuable outcomes are for intended beneficiaries. We further narrow our focus to one domain of the explicated conceptual framework, specifically the assessment skill domain. Within this narrowed context, in the next section we map out a description of important terms and the duties of an evaluation and of an assessment expert, highlighting where the duties of the evaluator and the duties of the assessment expert diverge and converge, and where, in theory, they can complement one another.

Mapping the Duties of the Assessment Expert

Before mapping the duties of the evaluator and the assessor, we summarize why assessor was selected to be juxtaposed against an evaluator. We also provide background on the term assessment and assessment expert, as understanding both are central to the argument contained herein.

The field of assessment and those who apply it, assessment experts, was selected because Alkin has noted that "assessment is a term that is often used synonymously with evaluation, but it is different" (2011, p. 10). Further, our own experience mirrors Alkin's claim. Each of us has expertise in evaluation and assessment, respectively, yet it is not uncommon for stakeholders, and even other faculty members at our institution, to perceive us as interchangeable experts on grants, contracts, and other applied projects. It is for this reason that evaluators and assessment experts are compared and contrasted.

Assessment is defined as applied measurement, i.e., the process of assigning a numerical value for some unobservable trait (Hambleton, Swaminathan, & Rogers, 1991). For example, in education, assessments are used in order to monitor students' progress and provide grades (Alkin, 2011). The assessment provides a tool to be able to assign a numerical value to a trait such as academic achievement. In psychology, assessments are used to observe human behavior as a way of measuring unobservable psychological traits such as intelligence, depression, aptitude, or extroversion (Furr & Bacharach, 2014). To illustrate, in psychology, assessments are used to observe human behavior as a way of measuring unobservable psychological traits such as intelligence, depression, aptitude, or extroversion (Furr & Bacharach, 2014).

The conversation around the term assessment begs the question, what is an assessment expert? An *assessment expert*, or *assessor*, is an individual whose job is to create and critically evaluate the instruments, data collection methods and protocols, checklists, and tests, in terms of their validity to the context, the reliability of the scores, and absence of bias.

Turning now to the mapping out the duties of an evaluator and an assessment expert, we return to the adapted Scriven (1994, 2015) framework set forth in the previous section, and use it to operationalize and understand the duties of the assessment expert. *Subject matter knowledge* needs to be current, comprehensive, and accurate. For the assessment expert, this would include having a solid foundational knowledge on assessment, including assessment theory, methods, and uses. Understanding the implications of the potential uses and misuses of the assessment is also a vital part of showing evidence of basic assessment skills and subject matter.

Application skill for assessors includes competence in developing and testing assessments to ensure the interpretations of the results are accurate and indicate evidence of validity. Before moving forward, it is important to understand terms such as validity, reliability, and absence of bias within the context of assessment. *Validity* refers to the degree to which evidence and theory support the interpretations of the assessment for the proposed purpose of the assessment (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). *Reliability* refers to the consistency in the scores of a test across replications (American Educational Research Association, et al., 2014). *Bias* refers to inherent issues with the assessment (either individual items or the construct in general) that would cause problems in the interpretability of the scores. For example, a self-report measure on personal eating habits may invoke respondents to answer questions in a way they may perceive to be correct, versus what is in fact the truth. Additionally, items on an assessment may favor a certain group of people inadvertently causing issues with the interpretations. In a large-scale assessment context, being able to select the appropriate measurement framework (e.g. Classical Test Theory versus Item Response Theory) is a necessary and fundamental skill. Additionally, assessment experts should be keeping up to date with the latest analytical methods such as new item response theory models (Hambleton, Swaminathan, & Rogers, 1991). Also, using an interpretive inquiry-based approach to gathering validity evidence for assessments (Deluca, 2010) can assist the assessment expert in being an informed developer and consumer of assessments.

Assessment skills are vital to the assessment expert. This category not only includes general knowledge on the assessment development process, but also includes setting standards, norming, linking and equating assessments and communicating the results, for example, to the evaluation team. An assessment expert is an individual whose job is to create and critically evaluate the instruments, data collection methods and protocols, checklists, and tests, in terms of their validity to the context, the reliability of the scores, and absence of bias. The process of designing an instrument is cyclical and iterative and begins with identifying conceptual and operational definitions of the constructs of the instrument. The operational definitions help to create measurable items, and identify appropriate response scales. Once the measurable items have been identified, the instrument will generally undergo a formal content review with substantive experts. These experts will review the items for their theoretical accuracy and evaluate their utility. Analysis of this data then leads to edits, revisions, and perhaps pre-piloting before a final version is formally piloted. Once piloted, various statistical tests will be conducted and provide a rationale for further edits, revisions, and additional analyses before a final version of the instrument is ready to use (Mccoach, Gable, & Madura, 2013). Assessment experts can use their knowledge of measurement to ensure large constructs are decomposed into observable items. Finally, assessment experts should have a deep understanding of some of the sophisticated analytic techniques, both qualitative and quantitative. For example, in a large-scale assessment context, assessment experts should be able to not only ensure the items are measureable, but be able to conduct systematic cognitive interviews and expert reviews (Willis, 1999), as well as examining measurement invariance and differential item functioning (Swaminathan & Rogers, 1990).

Professionalism for assessors touches on areas such as professional ethics, for example, maintaining the integrity of the assessments, ensuring the evaluation design does not include any assessment fraud, and following appropriate guidelines as necessary from the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014). The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education co-authored the *Standards for Educational and Psychological Testing*. "The purpose of the Standards is to provide criteria for the development and evaluation of tests and testing practices and to provide guidelines for assessing the validity of interpretations of test scores for the intended test users" (2014, p. 1)[5]. *The Standards for Educational and Psychological Testing* are organized into three broad categories: (i) Foundations, (ii) Operations, and (iii) Testing Applications. *Foundations* include standards regarding validity, reliability and fairness. *Operations* include standards related to administering scoring, and ethical responsibilities for the users. Finally, *testing applications* covers

[5] The *Standards for Educational and Psychological Testing* are written to refer to tests. We extend upon their focus to include the entire field of assessment across multiple disciplines, not just education and psychology.

standards for testing in various settings, including psychology, education, program evaluation, work place and credentialing, and policy studies. The *Standards for Educational and Psychological Testing* are not meant to serve as laws or rules, but as generic guidelines to ensure ethical practice with regard to administering, and making appropriate interpretations.

Implications for Evaluation

In this paper we have laid out one of the major ongoing tasks of Evaluation as a discipline and profession – the need to clearly communicate what is distinctive about evaluation in ways that resonate with stakeholders and other professionals. We have tied this work to larger conversations about how to maintain Evaluation's professional status, and reinforce our disciplinary walls. We have argued that one way to address this ongoing task is to develop a conceptual tool that can be used in dialogue with others, and have presented one such tool. We then offer a case example of how this conceptual tool could be used by mapping the knowledge and skills needed in evaluation compared to assessment, a related profession that is often confused with evaluation. The resulting mapping is depicted in Table 1.

While we have selected to compare evaluation and assessment, this conceptual tool can be extended by evaluation practitioners in a variety of ways. Evaluators engaged in evaluation capacity development work with stakeholders could use it in the early conversations with stakeholders to illuminate what is distinctive about evaluation work. Moreover, instead of or in addition to using assessment as a comparative case, practitioners could elect to extend the framework to other professions or disciplines. For example, if an evaluation client has a history of working with auditors, organizational psychologists, etc., a comparison between evaluation and the work of those professionals could be used and would be more appropriate to use.

Regardless of the comparison cases used, this kind of conceptual tool with comparative case examples would clarify for outsiders the distinct work of evaluators. Knowledge advancements in the science of learning across disciplines (c.f., Bransford, Brown, & Cocking, 2000) have generated a significant body of scholarship that suggests learners construct new knowledge and develop understanding drawing from what they already know and believe to be true. Thus, the job of any educator, including evaluators seeking to educate others, is to surface this prior knowledge and build upon it. It is through this process that learners re-examine and modify their prior understanding in light of new ideas. Moreover, there exists significant bodies of literature within and outside of evaluation that suggest it is not just about what is presented, but also how it is presented. The conceptual tool included in this paper allows for integration of prior stakeholder experiences because it is flexible and adaptable to the comparative cases selected, and also because it presents information in a way that is easily understandable to stakeholders. It is for these reasons that we believe this conceptual tool with comparative case examples would clarify what is distinctive about what evaluators know and can do.

This conceptual framework could also be used by others in evaluation for different reasons. To illustrate, it could be used by universities as a teaching tool to help novice evaluators understand differences between evaluation and content students are learning in other courses (e.g., statistics). It could be used by professional development providers to help development professionals and other types of would-be evaluators situate their existing knowledge within the discipline and profession of evaluation. It could also be used by researchers who study evaluation as the object of investigation (i.e., researchers engaged in research on evaluation) who want to study the use-in-practice of this conceptual tool or who want to generate evidence to inform evaluation capacity building efforts. It could be used by professional evaluators to map out what skills are needed for a particular evaluation study, or to inform why a particular type of expertise is needed on the evaluation team (e.g., assessment).

This paper presents initial efforts to develop a conceptual tool that could be used in dialogue with others about what is distinctive about evaluation. There is more work to be done, but by presenting initial efforts to develop this tool including foundational arguments about why it is important to do so, we hope to spark excitement about and work on communication tools that can address the need for greater clarity about the work of evaluation practitioners. We call on evaluation practitioners and scholars to take up work in this important area by using it in their practice, extending the conceptual tool, studying its use, or developing other evaluation communication tools.

Table 1

Comparing the Duties of Evaluation and Assessment Practitioners

Duties	Evaluation Practitioner Example	Assessment Practitioner Example
Generic Duties	▪ Subject matter knowledge: Knowledge about and ability to implement different evaluation theories or approaches ▪ Application skill: Knowledge about and ability to develop and execute a project management plan ▪ Assessment skill: Knowledge about and ability to be critical consumers of validity and reliability evidence for existing instruments ▪ Professionalism: Knowledge about and ability to use the *AEA Guiding Principles for Evaluators* ▪ Other secondary duties: Engaging with and drawing from the evolving evaluation knowledge base	▪ Subject matter knowledge: Knowledge about and ability to implement different assessment theories ▪ Application skill: Knowledge about and ability to develop and execute a project management plan ▪ Assessment skill: Knowledge about and ability to implement different standard setting procedures ▪ Professionalism: Knowledge about and ability to use the *Standards for Educational and Psychological Testing* ▪ Other secondary duties: Engaging with and drawing from the evolving assessment and measurement knowledge base
Job-Dpecific Duties	▪ In the context of an outcome evaluation: ▪ Subject matter knowledge: Knowledge about and ability to implement the specific evaluation theory or approach guiding the evaluation (e.g., Donaldson's theory driven evaluation science approach) ▪ Application skill: Knowledge about and ability to present information in ways that are perceived as credible, valid, and actionable by primary intended users ▪ Assessment skill: Knowledge about and ability to be a critical consumer of validity and reliability evidence related to outcome measures to be used (e.g., Peabody Picture Vocabulary Test, Fourth Edition) ▪ Professionalism: Knowledge about and ability to engage with local norms (e.g., community-specific IRB review boards) ▪ Other secondary duties: If using random assignment, meeting with different stakeholder groups to explain the process and hear concerns	▪ In the context of an outcome evaluation in which an assessment needs to be developed and validated: ▪ Subject matter knowledge: Knowledge about and ability to implement the specific assessment theory guiding assessment development (e.g., argument-based assessment theory) ▪ Application skill: Knowledge about and ability to present information in ways that are perceived as reliable, valid, and fair by both technical and non-technical audiences ▪ Assessment skill: Knowledge about and ability to examine how different items on the assessment are functioning ▪ Professionalism: Knowledge about and ability to engage with local norms (e.g., community-specific IRB review boards) ▪ Other secondary duties: Meeting with different stakeholder groups to explain the standard setting process and hear concerns
Context-Specific Duties	▪ Identification of the criteria to be used in meta-evaluative efforts of a particular evaluation in a particular context	▪ Identification of the criteria to be used in evaluative efforts of a particular assessment for a particular use

References

Alkin, M. C. (2004). Context adapted utilization: A personal journey. In M. Alkin, *Evaluation roots: Tracing theorists' views and influences* (pp. 293–303). Thousand Oaks, CA: Sage.

Alkin, M. C. (2011). Evaluation Essentials: From A to Z. Guilford Press.

Alkin, M. C. (2013). *Evaluation roots: A wider perspective of theorists views and influences* (2nd ed.). Thousand Oaks, CA: Sage.

Altschuld, J. W., & Engle, M. (2015). Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. *New Directions for Evaluation, 145*, 1–113.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

ANZEA & SuPERU (2015). *Evaluation Standards for Aotearo New Zealand*. The Aotearoa New Zealand Evaluation Association. Retreived

from http://www.anzea.org.nz/evaluation/evaluation-standards/

Blome, J. (2009). *National Institute of General Medical Sciences*.

Chen, H. T. (2015) *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2nd ed.). Thousand Oaks, CA: Sage.

Davidson, E. J. (2007). Marketing evaluation as a profession and a discipline. *Journal of MultiDisciplinary Evaluation, 2*, 3–10.

Davidson, E. J. (2015). Question-driven methods or method-driven questions? How we limit what we learn by limiting what we ask. *Journal of MultiDisciplinary Evaluation, 11*(24), i–x.

Davies, R., Randall, D., & West, R. E. (2015). Using open badges to certify practicing evaluators. *American Journal of Evaluation, 36*, 151–163. doi:10.1177/1098214014565505

Dewey, J., Montrosse, B. E., Schröter, D. C., Sullins, C., & Mattox, J. R. (2008). Evaluator competencies: What's taught versus what's sought. *American Journal of Evaluation, 29*, 268–287. doi: 10.1177/1098214008321152

DeLuca, C. (2011). Interpretive validity theory: Mapping a methodology for validating educational assessments. *Journal of Educational Research, 53*(3), 303-320. doi:10.1080/00131881.2011.598659

Donaldson, S. I. (Ed.). (2013). *The future of evaluation in society: A tribute to Michael Scriven*. Charlotte, NC: Information Age Publishing.

Dynarski, M., Rui, N., Webber, A., & Gutmann, B. (2017). Evaluation of the DC Opportunity Scholarship Program: Impacts after one year. NCEE 2017-4022. *National Center for Education Evaluation and Regional Assistance*.

Fitzpatrick, J. (2002). Dialogue with David Fetterman. *American Journal of Evaluation, 21*(2), 242–259. doi:10.1177/109821400002100214

Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation*, 68, 15–32.

Fournier, D. M. (2005). Logic of evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (p. 140). Thousand Oaks, CA: SAGE.

Freidson, E. (2001). *Professionalism, the third logic: On the practice of knowledge*. Chicago, IL: University of Chicago Press.

Furr, R., & Bacharach, V. (2014). *Psychometrics: An introduction* (Second ed.). Los Angeles, CA: Sage.

Gauthier, B., Kishchuk, N., Borys, S., & Roy, S. N. (2015). The CES professional designation program: Views from members. *Canadian Journal of Program Evaluation, 29*(3), 98–133. doi:10.3138/cjpe.29.3.98

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

House, E. R. (1977). *The logic of evaluative argument* (CSE Monograph 7). Los Angeles: University of California, Los Angeles, Center for the Study of Evaluation.

House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: SAGE. Reissued by Information Age Publications, NC, 2008.

House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. Newberry Park, CA: Sage.

Jacob, S. (2008). Cross-disciplinarization: A new talisman for evaluation? *American Journal of Evaluation, 29*(2), 175-194. doi:10.1177/1098214008316655

Jacob, S., & Boisvert, Y. (2010). To be or not to be a profession: Pros, cons and challenges for evaluation. *Evaluation, 16*(4), 349–369. doi:10.1177/1356389010380001

King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of MultiDisciplinary Evaluation, 9*(21), 11–21.

King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. Thousand Oaks, CA: Sage. doi:10.4135/9781452269979

King, J. A., & Stevahn, L. (2015). Competencies for program evaluators in light of adaptive action: What? So what? Now what? In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. New Directions for Evaluation, 145*, 21–37. doi:10.1002/ev.20109

Larson, M. S. (1977). *The rise of professionalism: A sociological approach* (Vol. 233). Berkeley, CA: University of California Press. doi:9780520029385

LaVelle, J. M. (2014). *An examination of evaluation education programs and evaluator skills across the world*. (Unpublished doctoral dissertation). The Claremont Graduate University, Claremont, CA.

LaVelle, J. M., & Donaldson, S. I. (2010). University-based evaluation training programs in the United States 1980–2008: An empirical examination. *American Journal of Evaluation, 31*(1), 9–23. doi:10.1002/ev.20110

LaVelle, J. M., & Donaldson, S. I. (2015). The state of preparing evaluators. In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. New Directions for Evaluation*, *145*, 39–52. doi:10.1002/ev.20110

Lincoln, Y. S. (1985). The ERS standards for program evaluation: Guidance for a fledging profession. *Evaluation and Program Planning, 8*, 251–253. doi:10.1016/01497189(85)90046-1

MacDonald, K.M. (1995) *The Sociology of the Professions*. Thousand Oaks, CA: Sage. doi:10.1177/0730888488015003001

Mathison, S. (2007). What is the difference between evaluation and research—and why do we care? In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 183–196). New York, NY: Guilford.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: Social and corporate applications* (3rd ed.). New York: Springer. doi:10.1007/978-1-4614-7135-6

Millerson G (1964). *The qualifying associations: A study in professionalization*. London: Routledge and Kegan Paul Ltd.

Nilsson, N., & Hogben, D. (1983). Metaevaluation. In E. R. House (Ed.), *New directions for program evaluation: No. 19. Philosophy of evaluation* (pp. 83–97). San Francisco, CA: Jossey-Bass.

Nunns, H., & Roorda, M. (2010, May). *Final evaluation report of the recognized seasonal employer policy (2007-2009)*. New Zealand: Evalue Research team, Research New Zealand, the University of Auckland, and the Department of Labour. Retrieved from http://thehub.superu.govt.nz/publication/final-evaluation-report-recognised-seasonal-employer-policy-pdf

OECD (2002). Glossary of key terms in evaluation and results-based management. Paris, France: OECD/DAC.

Oksanen, R. (2016, October). *Professionalization of evaluation in Europe*. Paper presented at the meeting of the American Evaluation Association, Atlanta, GA.

Patel, M. (2013, African Evaluation Guidelines. *African Evaluation Journal, 1*, 15. . http://dx.doi.org/10.4102/aej.v1i1.51

Patton, M. Q. (1990). The challenge of being a profession. *Evaluation Practice, 11*(1), 45–51.

Perrin, B. (2005). How can information about the competencies required for evaluation be useful? *Canadian Journal of Program Evaluation, 20*(2), 169-188. doi:10.1016/0886-1633(90)90008-2

Picciotto, R. (2011). The logic of evaluation professionalism. *Evaluation, 17*(2), 165–180. doi:10.1177/1356389011403362

Picciotto, R. (2010, January 13). Country-led M&E systems -Robert Picciotto part 2 [Video file]. Retrieved from http://www.youtube.com/watch?v=UfNakPTODBs%3E.

Porter, S., & Goldman, I. (2013). A growing demand for monitoring and evaluation in Africa. *African Evaluation Journal, 1*(1). doi:10.4102/aej.v1i1.25

Profession. (2017). In OxfordDictionaries.com. Retrieved from https://en.oxforddictionaries.com/definition/profession

Professionalize. (2017). In OxfordDictionaries.com. Retrieved from https://en.oxforddictionaries.com/definition/professionalize

Rodríguez Bilella, P. D., Martinic Valencia, S., Soberón Alvarez, L., Klier, S. D., Guzmán Hernández, A. L. & Tapella, E. (2016) (2016). *Evaluation Standards for Latin America and the Caribbean* (1st ed.). Ciudad Autónoma de Buenos Aires. Retrieved from https://www.deval.org/en/105/evaluation-standards-for-latin-america-and-the-caribbean-published.html

Rog, D. J., Fitzpatrick, J. L., & Conner, R. F. (2012) Context: A framework for its influence on evaluation practice. *New Directions for Evaluation, 135*, 1–105.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage. doi:10.1097/00004356-198009000-00034

Schwandt, T. A. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Stanford, CA: Stanford University Press.

Scriven M. (1975). *Evaluation bias and its control.* Occasional Paper Series sponsored by the National Science Center (Paper #4). Kalamazoo, MI: Evaluation Center, Western Michigan University.

Scriven, M. (1991). *Evaluation thesaurus* (Fourth ed.). Thousand Oaks, CA: SAGE.

Scriven, M. (1993). The nature of evaluation. In M. Scriven (Ed.), *New directions for program*

evaluation: No. 58. Hard-won lessons in program evaluation (pp. 5–48). San Francisco, CA: Jossey-Bass. doi:10.1002/ev.1640

Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education, 8*, 2, 151–184. doi:10.1007/BF00972261

Scriven, M. (2007). The logic of evaluation. In H. V. Hansen, D. M. Godden, H. V. Hansen, C.W. Tindale, & J. A. Blair (Eds.), *Dissensus and the search for common ground* [CD]. Windsor, ON: Ontario Society for the Study of Argumentation (OSSA).

Scriven, M. (2015, February). Duties of the teacher [updated]. Retrieved from http://michaelscriven.info/papersandpublications.html

Scriven, M. (2015, August). Key evaluation checklist [updated]. Retrieved from http://michaelscriven.info/papersandpublications.html

Segone, M. (2010). *From policies to results: Developing capacities for country monitoring and evaluation systems*. Retrieved from: http://mymande.org/sites/default/files/From_Policy_To_Results.zip

Shadish, W. R. (1999). Evaluation theory is who we are. *American Journal of Evaluation, 19*(1), 1–19. doi:10.1016/S1098-2140(99)80177-5

Stufflebeam, D. L. (1978). Meta evaluation: An overview. *Evaluation and the Health Professions, 1*, 17–43. doi:10.1177/016327877800100102

Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation, 22*, 183–209. doi: 10.1016/s1098-2140(01)00127-8

Stufflebeam, D. L. (2011). Meta-evaluation. *Journal of MultiDisciplinary Evaluation, 7*, 99–158. (Original work published 1974).

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, & applications* (2nd ed.). San Francisco, CA: Jossey-Bass.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361–370. doi:10.1111/j.1745-3984.1990.tb00754.x

Toulmin, S. E. (1972). Human understanding: The collective use and evolution of concepts (Vol. One) Princeton, NJ: Princeton University Press.

Willis, G. B. (1999, August). *Cognitive interviewing: A "how to" guide.* Paper presented at the meeting of the American Statistical Association, Baltimore, MD. doi:10.4135/9781412983655