EJA | EVALUATION
Journal of Australasia

Moving from guideline to measure to findings: The Australian Evaluation Society and the Learn Evaluation Assessment Portal

Amy M Gullickson ⓘ
The University of Melbourne, Australia

Taimur Siddiqi
Victoria Legal Services Board and Commissioner, Australia

Delyth Lloyd ⓘ
Victoria Department of Health, Australia

Anne Stephens ⓘ
Ethos of Engagement, Australia

George Argyrous
Paul Ramsay Foundation, Australia

Lauren Wildschut ⓘ
Stellenbosch University, South Africa

Abstract

In 2013, the Australasian Evaluation Society (AES) launched the Evaluators' Professional Learning Competency Framework. In 2020, the AES (now Australian Evaluation Society) partnered with learnevaluation.org to develop an online self-assessment for the AES community. The AES Competencies were explicitly designed to support professional learning, not as any kind of assessment instrument. Therefore, work was

Corresponding author:
Amy M Gullickson, The University of Melbourne, Melbourne, VIC 3010, Australia.
Email: amy.gullickson@unimelb.edu.au

needed to update the competencies into a measurable format. The authors report the history of the competencies, developments made before and after pilot testing them in the online Learn Evaluation Assessment Portal (LEAP), and the findings from the data collected over two and a half years of using the LEAP. The AES Competencies Framework remains one of the few competency sets with explicit knowledge and skill statements related to the logic of evaluation.

Keywords
evaluator competencies, evaluator self-assessment, evaluation capacity building, evaluation professional associations, research on evaluation

---

What we already know

- Evaluator competency sets have become a way to define the profession of evaluation.
- Self-assessments are useful for planning personal and organisational evaluation capacity building.
- Few self-assessments of evaluator competencies exist.

The original contribution the article makes to theory and/or practice

- History of the AES Evaluators' Professional Learning Competency Framework.
- Publication of updated AES competency framework used in the self-assessment.
- Analysis of findings on the competencies from the self-assessment.

Introduction

For evaluation to fully mature as a profession, there needs to be a process for determining the necessary knowledge, skills, and values which evaluators require – no matter where they practise (Smith, 1999; Worthen, 1994; Worthen & Sanders, 1991). Evaluator competencies, 'the essential knowledge, skills, and dispositions that evaluators need to conduct program evaluations effectively' (Ghere et al., 2006, p. 109) have been developed by academics, Voluntary Organisations for Professional Evaluation, and other organisations worldwide to address this need in their contexts.

Organisations with evaluator competency frameworks use them for a range of purposes, including credentialing evaluators (e.g., Canadian Evaluation Society and Japanese Evaluation Society), and focussing professional development (Stevahn et al., 2005). King and Stevahn (2020) argued that the five domains of evaluator competency are largely agreed upon internationally. However, explicitly evaluative competencies related to valuing and the logic of evaluation are often not included, or not explicit – and these are core for a profession where its purpose includes value judgements.

The Australian Evaluation Society (AES) has been engaged with evaluator competencies since 2012. The AES Evaluators' Professional Learning Competencies Framework features two domains and several competencies that explicitly cover evaluation-specific skills, knowledge, and attitudes. Since 2020, members of the AES Pathways Committee and the researchers from learnevaluation.org have been working collaboratively on advancing evaluation learning through self-assessment using the AES evaluator competencies. This article presents that work as a part of the larger narrative on evaluator competencies, discusses the development and evolution of the AES Framework, and reports on findings from several years of self-assessment data.

History of the AES Evaluators' Professional Learning Competencies Framework

In 2012, the Australasian Evaluation Society (now Australian Evaluation Society) Board charged its Professional Learning Committee to undertake creation of an evaluator competency framework for the organisation. The task fell within the Committee's remit to provide 'leadership and guidance on professional learning matters including programs and other activities offered by the AES' (AES Professional Learning Committee, 2013, p. 3). The Committee comprised AES members from Australia (11) and New Zealand (1), including the Executive Officer and Professional Learning Coordinator of the AES. Committee members included practitioners (7) working in government, not-for-profit and for-profit organisations, and academics (2). Two practitioners were also PhD students at that time. The Board mandated that the competency framework should 'guide and support members and other interested parties to enhance their evaluation knowledge and expertise' (p. 3), not measure skills and knowledge for credentialling or professionalisation purposes. The risk of competencies serving as a tool for exclusion from the profession had impeded previous efforts.

At that time, seven evaluator competency frameworks had been published: three from the United States (Russ-Eft, 2008; Stevahn et al., 2005; Visitor Studies Association, 2008), and one each from Canada (Canadian Evaluation Society, 2010), the Pacific (Aotearoa New Zealand Evaluation Association, 2011), Europe (European Evaluation Society, 2012), and international development (IDEAS Working Group on Evaluator Competencies, 2012). The Committee chose four of these as relevant to their work. The right-hand columns in Table 1 present these sets and their domain structures: Aotearoa New Zealand Evaluation Association (anzea), the Canadian Evaluation Society, the European Evaluation Society, and the International

Table 1. Comparison of competency domains.

Australian Evaluation Society (2013)	Aotearoa New Zealand Evaluation Association (2011)	Canadian Evaluation Society (2010)	European Evaluation Society (2012)	International Development Evaluation Association (2012)
Evaluation Attitude and Professional Practice	Reflective practice and professional development	Reflective practice	Dispositions and attitudes	Professional Foundations
Theoretical Foundations	Systematic evaluative inquiry	Technical practice (no evaluation-specific methods)	Evaluation knowledge: 1 and 2	Monitoring systems
Research Methods and Systematic Inquiry			Evaluation knowledge: 3	Evaluation Planning and Design
Attention to Culture, Stakeholders, and Context	Contextual analysis and engagement Values and Cultural Competency[a]	Situational practice (no explicit attention to culture)		Managing the evaluation
Project Management	Evaluation project management and professional evaluation practice	Management practice	Professional practice 1	Conducting the evaluation
Interpersonal Skills	Evaluation project management and professional evaluation practice	Interpersonal practice	Professional practice 2	Communicating evaluation findings Promoting a culture of learning from evaluation
Evaluation Activities				

[a]This Domain is listed first and serves to ground all others in the anzea framework.

Development Evaluation Association. The Essential Competencies for Program Evaluators (King et al., 2001; Stevahn et al., 2005) were the basis for the Canadian competencies, so the Committee did not include them separately. The Committee mapped the domains and competencies across the four sets, debated, and devised their

own version of the domains and assignation of competencies (Table 1, far left column). The far right column of Table 1 is separated from the rest by a line to indicate that the domains in the International Development Evaluation Association were based on the life cycle of an evaluation; competencies within their domains map across domains from the other sets.

Figure 1 presents the Committee's vision of the relationship among the AES competencies. Evaluators' knowledge and skills are grounded in Evaluation Attitude and Professional Practice (Domain 1). Evaluation as a task involves bringing together values and facts in particular contexts to create judgements, displayed by the relationship between Theoretical Foundations (Domain 2), Research Methods and Systematic Inquiry (Domain 3), and Culture Stakeholders and Context (Domain 4). Project Management (Domain 5) and Interpersonal Skills (Domain 6) provide the logistics and people skills necessary to conduct evaluations. Those domains all culminate in Evaluation Activities (Domain 7) – the actual delivery of evaluation. Key issues that generated the domains and grounding of the Framework are discussed below.
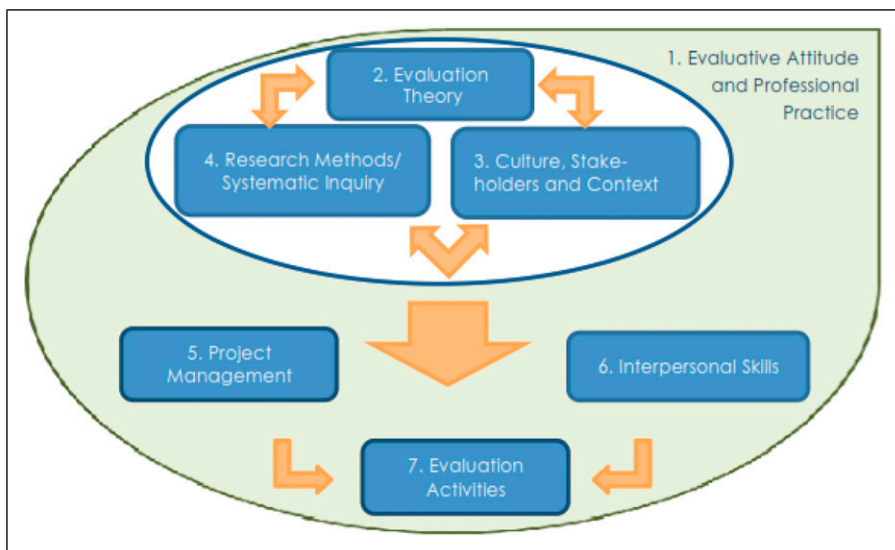
Figure 1. The Inter-relationship of the AES Evaluators' Professional Learning Competencies. (AES Professional Learning Committee, 2013, used by permission).

The first key issue generated the Theoretical Foundations domain. Several members of the Committee expressed concern that the existing competency sets were not explicit enough about evaluation-specific knowledge, skills, and methods, and the value-based reasoning needed for good practice. A review of literature surfaced a potential explanation for the situation. In 2005, Stevahn, King, Ghere, and Minnema reported they

had intentionally removed critical thinking from their original competency set because it was implicit in the practice of evaluation:

> Finally, we also omitted the 'logical and critical thinking' competency (IIIA) in King et al. because such thinking is an integral component underlying virtually all other evaluator competencies, especially those that directly require determining strengths and limitations of studies, providing rationales for decisions, weighing alternatives, making reasoned judgements, and so on (see, e.g. Table 1, competencies 2.17 and 2.19). (p. 54)

As those authors' Essential Competencies for Program Evaluators (2001 and 2005) pre-dated and served as templates for many other competency sets, it seems likely this omission was disseminated.

In response, the Committee used a quote from Shadish (1998) to justify the inclusion of the Theoretical Foundations domain:

> And if you do not know much about evaluation theory, you are not an evaluator. You may be a great methodologist, a wonderful philosopher, or a very effective program manager. But you are not an evaluator. To be an evaluator, you need to know that knowledge base that makes the field unique. That knowledge base is evaluation theory… (pp. 5–6)

Within that domain, the committee intentionally included competencies related to the language of evaluation, the purpose of evaluation, the logic of evaluation, evaluative actions, synthesis methodologies, and ethical and moral reasoning. A final differentiator was the addition of a competency on cost analysis and value for money, which was not explicitly included in any other competency set at the time.

The Committee also made choices informed by the content and structure differences present in the anzea and International Development Evaluation Association frameworks, manifested in the domains Evaluation Attitude and Professional Practice; Attention to Culture, Stakeholders, and Context; and Evaluation Activities. The former two domains arose from an examination of the anzea competencies (Aotearoa New Zealand Evaluation Association, 2011) which were grounded in values and cultural competency. This was unique for two reasons. First, in 2012, cultural competency had no or limited mention in competency frameworks. The committee agreed that competencies related to culture were essential and applied across all domains. Like anzea, the Committee chose to create a domain specifically for those competencies [Attention to Culture, Stakeholders, and Context], bringing together situational practice and context with culture. The second unique feature was anzea's grounding of their competencies in a specific domain. The Committee adopted the idea of grounding, choosing the domain Evaluation Attitude and Professional Practice because '…self-reflection and ongoing professional development…are critical in the broad and complex role of evaluators. This set of knowledge, skills, and attitudes influence all the other competency groups' (AES Professional Learning Committee, 2013, p. 7). The Evaluation Activities domain, a culminating list of tasks the evaluator would carry out

in an evaluation, was included because the Committee appreciated the organisation of the International Development Evaluation Association framework using the lifecycle of an evaluation (AES Professional Learning Committee, 2013, p.7).

AES shared the draft competencies with members in early 2013 for feedback. People provided feedback via email and personal conversations to committee members. One of the major discussions was the definition of the purpose of evaluation as 'making judgements of merit, worth, and significance' (AES Professional Learning Committee, 2013, p. 11). The committee debated and discussed this at length but decided to leave the statement unchanged. Other comments were addressed, and changes made where the committee agreed. The AES Board endorsed the Framework in July 2013 (A. M. Gullickson, personal communication from CEO), and it was launched at the AES conference in Brisbane in September 2013.

Validation study

In 2015, a Master of Evaluation student from the University of Melbourne Centre for Program Evaluation conducted a validation study of the Framework (Putra, 2015) under the auspices of the Committee. The student conducted an online survey of all AES current and past members. Twenty-eight people responded: 18 current members and ten non-members: '…respondents were from Australia (20), Philippines (1), Indonesia (3), Timor-Leste (1), Congo (1), New Caledonia (1), and Canada (1)' (Putra, 2015, p. 10). They ranged in experience with evaluation from 1 to 2 years to more than 15 years, with the majority working more than 15 years (n = 12). Four respondents involved in developing the framework responded to the survey.

The survey used 11-point scales with top and bottom anchors (0 very unlikely– 10 very likely; 0 definitely unconfident–10 highly confident). Respondents reported they were likely to use the Framework for developing professional competency ($n = 18$, mean = 7.8, sd = 1.87) and relatively confident in it as a guide for professional competency ($n = 18$, mean = 7.94, sd = 1.44). The four respondents who had been involved in developing the framework reported a higher level of confidence and use.

As a result of the validation study and other ongoing work, the Committee drafted an unpublished plan that called for the AES to use the competencies to identify and address gaps in their workshop programs, create a mentoring program, and develop a self-assessment tool (A. M. Gullickson, personal records). In 2019–2020, the Committee was renamed the Pathways Committee (Australian Evaluation Society, 2021). The Committee established a Working Group to create a self-assessment for the evaluator competencies. The remainder of this article discusses the two phases of that process: (a) getting to a measure from the 2013 framework, and (b) testing and revisions of the self-assessment related to the competency statements.

Methodology and methods

When the Working Group undertook developing the self-assessment in 2020, they began from the original Framework. It had been explicitly designed as a guideline for self-reflection and learning, not for measurement. Thus, work was needed to update the competency statements for online measurement. This required attention to survey design (Dillman et al., 2014) and principles of assessment validity (Gliner et al., 2017; Holden, 2010). Good surveys have clear, single barrelled items with only intentional duplication, phrasing that enables assessment, and a clearly anchored rating scale. In contrast, competencies are statements that combine knowledge, performance, disposition, and reflection (Austin, 2019; Hodges, 2015). The breadth of frameworks for understanding competence and the breadth within competency statements made conversion a challenge.

In creating an assessment, the argument for instrument quality can be based on a continuum of evidence including face, content, criterion, and construct validity (Cook et al., 2015; Coolidge & Segal, 2010). For self-assessment of the evaluator competencies, criterion and construct validity were not possible as there were no established empirical measures of the competencies, and no agreed measures of evaluator job performance. As a result, the work described in this article focused on face and content validity with the aim of establishing a self-assessment that respondents felt was appropriate and relevant to their evaluation practice.

The factors for establishing face validity include: (a) the opinions of test takers, (b) the obviousness of the test item content, and (c) the situation in which the test is administered (Holden, 2010). The steps for establishing content validity include: (a) defining the concept that the test is attempting to measure, (b) searching the literature for additional definitions, (c) generating items, and (d) reducing the number of items through testing and expert review to assess representativeness (Gliner et al., 2017).

The Working Group tackled the steps for creating a valid self-assessment of competencies through two iterative phases. In Phase 1, they revised the competencies into assessment items. In Phase 2, the Working Group engaged with the researchers at learnevaluation.org and Matilda Tech to design an online portal for the self-assessment. Once the Learn Evaluation Assessment Portal (LEAP) was ready, the Working Group invited AES members and the public to engage with the self-assessment. This allowed the Working Group and the researchers to gather data to refine the assessment and to understand the demographics of, and evaluation expertise amongst, respondents in relation to the competencies. The research questions for the phases were as follows:

Phase 1: getting to a measure

   (1) What changes need to be made to the 2013 AES Evaluators' Professional Learning Competency Framework to convert it into a valid self-assessment instrument?

(2) Based on the pilot tests, what changes need to be made to increase face and content validity of the competency statements in the self-assessment?

Phase 2: data and findings

(1) Does the instrument have face and construct validity?
(2) What can be learnt about a sample of evaluators from a competencies self-assessment?

The two phases were iterative; the competency set was revised as follows:

(1) 2020: in preparation for the launch – the most significant revision, attending to duplication, clarity, and utility for measurement.
(2) 2021: After the alpha and beta test – minor adjustments based on feedback from alpha testers.
(3) 2022: Re-numbering the domains to align with the self-assessment format.

Throughout, revising the Framework was done by the Working Group; the re-searchers provided analysed feedback from the pilot test and adapted the LEAP accordingly with revisions. The first author bridged between the groups.

Methods

For Phase 1, in early 2020, the Working Group broke the Framework into subgroups by domain and analysed the competencies for clarity, utility for assessment, and dupli-cation, via a Word document. In late July, the revisions were consolidated into an Excel file and shared via Google docs. The Working Group re-assembled over three virtual meetings to discuss revisions proposed by the subgroups. During those meetings, the whole group went through the document and discussed each competency to ensure the current version captured anything potentially missed earlier. Because the Theoretical Foundations domain was unique to the Framework, a group member went back to the evaluation literature around the logic of evaluation to check the content in the com-petencies. The Working Group repeated the review and discussion process between versions, incorporating feedback from respondents into the revision, which led to additional changes and reductions in items.

For Phase 2, the revised Framework was loaded into the LEAP. Four separate survey links were created and distributed over 2020–2022. The instrument was distributed at AES events, and via a public website hosted by the research team. The research team's status as recognised evaluation academics helped establish face validity. Data from these survey links was collected from AES members and the general public. In total, across all distributions of the survey, there were 294 responses:

• Alpha test (2020) = 116 respondents.

- Beta test (2021) = 7 respondents.
- Gamma test (2021) = 58 respondents.
- Current version (2022) = 113 respondents.

Of those responses, 180 were either AES members or non-AES attendees at the FestEVAL events in 2020 and 2021. The remainder ($n = 114$) were respondents via the public portal. As a result, the population for the study is unclear; attendees at the AES events numbered in the hundreds and the public website made the LEAP available internationally. Thus, findings should be considered descriptive only and not generalisable to the whole population of AES or evaluators internationally.

Ethical considerations

The LEAP and learnevaluation.org are part of an international, multi-year research effort led by University of Mississippi. The original research was reviewed by their Institutional Review Board in 2019; the Portal was submitted as an additional data collection method in 2020. The review process '…determined [this project] does not meet the regulatory definition of human subjects' research at 45 CFR 46.102 and thus does not require Institutional Review Board approval'. However, the team was committed to ensuring clarity about the use, privacy, and confidentiality of the data participants shared.

Therefore, in Phase 2, participants were informed about use of the data in the live sessions. Following the alpha test, the learnevaluation.org team and Matilda Tech and included a privacy statement and retrospective consent. Alpha test data was only included in the analysis if the participants consented. For all other data collection rounds, respondents were presented with a consent page asking them to confirm they were over 18 and giving them the option of whether their de-identified data could be included in the research on evaluator competencies. Those who decline to have their data included still get a report, but their identifying information and responses are not saved in the LEAP database. For all consenting respondents, names and login details are kept only by Matilda Tech. All data are de-identified before any reports or data files are shared with participating organisations and the learnevaluation.org team. For more details about the ethical considerations related to the LEAP, see Wildschut et al. (2024, this issue). The following sections present findings for Phase 1 and Phase 2.

Phase 1: getting to a measure

Initial revisions

The original AES Framework had 111 competency statements across 7 domains:

(1) Evaluation Attitude and Professional Practice
(2) Theoretical Foundations

(3) Attention to Culture, Stakeholders, and Context
(4) Research Methods and Systematic Inquiry
(5) Project Management
(6) Interpersonal Skills
(7) Evaluation Activities

In 2020, the AES Pathways Working Group [the Group] began the updating process by addressing structural issues. First, they numbered the competencies to enable ease of reference. Second, the structure included domains, competencies, and in several cases, sub-competencies; this was too complicated. As the Group reviewed, they looked for ways to shift language and order of competencies within domains to remove the need for subsets, and adjusted as needed to ensure the competency statements could stand on their own as items within the domain.

In terms of content, the Group reviewed individual competency statements and identified issues with clarity, duplication, and utility for assessment:

- Forty-nine statements required updates for clarity. For example, with Original Competency 1.1.6: 'Practice meta-evaluation, seek ways to incorporate accountability into their work' the group agreed that wording created confusion because the link between meta-evaluation and accountability was not clear. The Group noted that US program evaluation standards (Yarbrough et al., 2010) have Evaluation Accountability Standards, which include statements on both internal (E2) and external meta-evaluation (E3). The Group agreed to change the competency statement to 'Seek ways to incorporate accountability into their work'. This shortened and clarified the statement.
- Ten items were flagged as potential duplicates. Some were merged to improve clarity; two competencies were removed:
  ○ 4.6 'Understand the range of methods available and the most appropriate mix of methods for the evaluation', which was identified as the overall theme of the Domain, and
  ○ 5.3 'Define work parameters, plans, and agreements', which was identified as overlapping with both 5.1 'Translate evaluation knowledge and theories into a workable plan for conducting the evaluation', and 5.2 'Scope an evaluation plan and negotiate a contract where relevant'.
- Four items were tagged with utility issues and two with problems of both duplication and utility. For example, 1.1.1 'demonstrate flexibility' was considered hard to measure but important, so an example was added to make it 'demonstrate flexibility, for example, able to respond and adjust to changing circumstances during an evaluation'.

Six competency statements were removed overall – the two mentioned above plus the following four:

- '1.1 Maintain integrity in their practice' was judged as too vague, and covered in two other competency statements.
- '1.2 Build their professional practice', which was a heading for a sub-section.
- '2.7 Ethical and moral reasoning (e.g, identifying and managing bias, coercion, or harm)' was removed as it is stated in the AES ethical guidelines and listed as a competency in Domain 1. (1.1 Know and uphold professional evaluation ethics).
- '3.9 Understand how evaluation relates to the other functions within the context' was removed due to being too broad and poorly defined, making it redundant.

Three new competencies were identified as missing and important, and therefore added:

- 3.17 Identify and access data sources.
- 4.08 Articulate the theory of change or logic model that will guide the evaluation and is tailored for the evaluation context.
- 5.12 Identify and mitigate problems and issues in constructive and useful ways (e.g. through giving and receiving feedback, de-escalating confrontation, and fair and even handed problem resolution).

Phase 1a: Testing yields further revisions

The updated competencies were loaded into the LEAP and pilot tested at FestEVAL 2020 (for more details about the alpha test, see Wildschut et al., 2024, this issue). 100 people took the self-assessment during that event and provided feedback live in the Zoom chat, an item at the end of the self-assessment, and via email. The process generated 231 comments related to specific questions in the tool. The research team categorised and summarised all the feedback by item so it could be addressed by either the research team or the Working Group. In revising the tool for future use, the research team handled feedback related to the technical issues, format in the LEAP, and demographic items. The Working Group handled competency related feedback and revised the expertise scale (Wildschut et al., 2024, this issue).

A variety of responses highlighted the range of roles evaluators take on the variation in relevance of competencies related to individual's roles:

'While I can, and do, do all these things I don't necessarily do them all in every evaluation. There needs to be an introduction that clarifies what we are expecting to see and how people should interpret the results. Should probably also be in the results'.

'Q14 made me realise that many of the tasks of project management are done by others (e.g. finance folks if you are in a uni, consultancy, private firm, or large [not-for-profit]) and so am I rating my personal skills at each of these things, or my ability to get the right people to do them the right way?'

Those and similar comments led the Group and the researchers to add an introductory sentence that no one person is expected to have expertise in all competencies. Another concern raised by one participant was the lack of a privacy statement. The research team apologised for that oversight and provided one, along with an additional consent for the following iterations of the LEAP. The Group and the researchers launched the updated version of the self-assessment in 2021.

Summary

In Phase 1, we posed Research Question 1: What changes need to be made to the 2013 AES Evaluators' Professional Learning Competency Framework to convert them into a self-assessment instrument? The answer was MANY!

To prepare for the alpha test, across six of the seven domains, only 35 competencies when checked had no concerns or changes. The Group shifted the order of 31 competencies within domains, updated the wording for 22 competency statements, added examples to 16, and definitions to eight. Six competency statements were removed and three added. The total number of competency statements went from 111 in the original to 95 for the alpha test. All of this was done within the bounds of the original set; the Group noted, but did not make, significant changes, as the original Framework had AES Board and Member approval.

To answer Research Question 2 (Based on the pilot tests, what changes need to be made to increase face and content validity of the competency statements in the self-assessment?) several additional changes were made. The Group added a statement to clarify that no evaluator should be expected to have expertise in all the competencies, revised the numbering of the domains, added additional examples, made some minor changes to order, and added a privacy statement. Changes were also made to the demographic items and the rating scale (see Wildschut et al., 2024, this issue). The Updated AES Evaluator Professional Learning Competency Framework (2022) Domain names and numbers are:

(1) Evaluation Activities
(2) Evaluation Attitude and Professional Practice
(3) Theoretical Foundations
(4) Attention to Culture, Stakeholders, and Context
(5) Research Methods and Systematic Inquiry
(6) Project Management
(7) Interpersonal Skills

The final, refined set of AES competencies now on offer in the LEAP is presented in Appendix A. The next section presents the data and findings from all the iterations (2020–2022).

Phase 2: Testing the measure

All the data collected through the LEAP have been analysed and presented in this section. The working group and research team cleaned and consolidated it to account for changes in the demographic and competency items across the versions of the survey. For clarity and simplicity, the LEAP generates individual and organisation level reports using arithmetic means for Domain ratings (Wildschut et al., 2024, this issue). Therefore, the authors used the arithmetic means to analyse the overall data set along with descriptive statistics to compare across cohorts.

Demographics of respondents

Nearly three-quarters of respondents were female (73% vs. 26% male, 1% other). Half of the respondents were between the ages of 35–49, with the rest roughly split around that range (25% under 35, 20% 50–59, and 9% 60 or above). For most, a master's degree was their highest qualification, followed by almost equal parts bachelor's and secondary. Respondents were able to select multiple disciplines. Of those who reported tertiary level study (78%), nearly half selected two or more disciplines. The most common were History and Human Society (33% of respondents), Sciences (20%), and Health and Medicine (19%). Participants reported practicing evaluation in Asia, the Pacific, Europe, North America, Africa, the Middle East, North America, and South America.

Almost 70% had completed some formal study in evaluation. Across all respondents, the median number of years in the workforce was 20, and the median number of years working in evaluation was eight. A plurality of respondents (43%) worked in government, while a further 20% and 18% worked in not-for-profit organisations and consultancies, respectively. The remaining 19% worked in for-profit organisations, academic institutions, or 'other'. Almost half (47%) of respondents reported that they had three-quarters to a full workload dedicated to evaluation; 25% reported that evaluation comprised less than a quarter of their role. When asked if they felt they had major gaps in their preparedness to work in evaluation, over 70% said yes.

To better understand the profile of respondents, the authors segmented and compared a sample of demographic variables. Those currently working in government reported much higher amounts of evaluation in their role, with 62% stating that it comprised 75% or more of their role, compared to 30%–40% for those working in other contexts. When comparing perceived gaps in preparedness to work in evaluation, differences were based on gender and years working in evaluation. A slightly larger proportion of female respondents reported major gaps, 76% versus 63% for men, as did those who had worked in evaluation for less than 10 years (80%) compared to those who had worked in evaluation for 10 or more years (59%). Comparing years working in evaluation and perceived major gaps in preparedness to work in evaluation revealed a weak negative correlation (r = −0.11).

Table 2. Competency Rating Scale adapted from Dreyfus and Dreyfus (2005).

| Rating	Description
None	I don't know anything about this domain or competency
Novice	I have heard about this domain/competency, but the vocabulary or concepts are unfamiliar and/or I don't really understand how to it do in practice
Advanced beginner	I have had some experience with this domain/competency either through examples or experience. I have a basic understanding of what it is and how it works, which I can apply in contexts where I have used it before. Outside of those contexts, I need help
Competent	In this domain/competency, I am able to analyse, calculate, and prioritise information to determine what needs to be done across contexts and situations. I make reasoned decisions in new situations based on my experience, but that takes effort and I'm less sure about the appropriateness of those choices
Proficient	In this domain/competency, I have a variety of positive and negative experiences that have shaped my perspective. I can see holistically what needs to be done in each situation, even when it varies from familiar patterns. But, it still takes some analytical thinking and planning
Expert	In this domain/competency, I operate on intuition honed over years of experience. I am able to see the total situation and know immediately how best to adapt my experience to practice effectively in different contexts. I only use analytical approaches when there's a problem or a novel situation

Respondents + the competencies

Respondents rated themselves on each competency using an expertise scale based on Dreyfus and Dreyfus (2005) presented in Table 2. For further discussion of the scale and its development, see Wildschut et al. (2024, this issue).

Self-assessment responses, by rating distribution

Within the data set, there were two groups (AES members and the public portal). Once the data set was consolidated, the working group looked at Domain level differences between groups. The AES members group did score consistently slightly higher on all Domains; t-tests showed statistically significant difference for most Domains. However, the differences between the groups disappeared when the data was filtered by years working in evaluation and percentage of evaluation in their current work. Since these were the main significant variables, the authors have chosen to present the whole data set focused on those, rather than the groups.

   Figure 2 shows that across all the competencies 67% of respondents self-assessed as Competent or above. Figure 3 shows the Domain level distribution of expertise ratings; most respondents self-assessed as Competent or higher in all Domains. This ranged
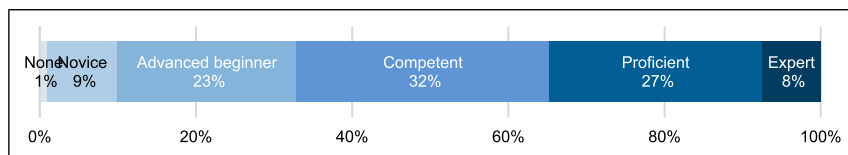
Figure 2. Distribution of competency scores, all competencies aggregated.
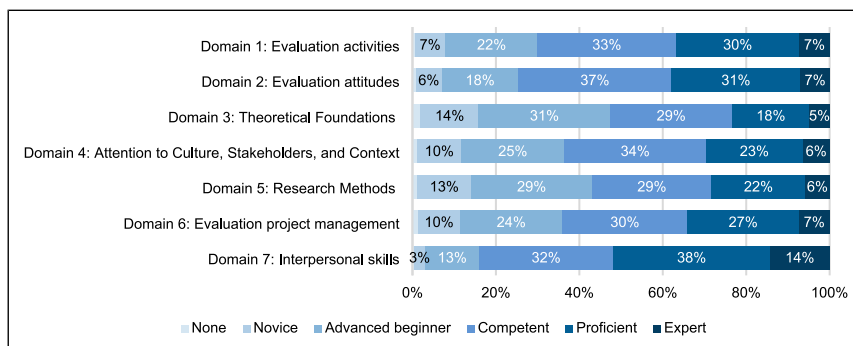
Figure 3. Distribution of competency scores, by Domain.

from 52% of all respondents for Theoretical Foundations (Domain 3) to 84% for Interpersonal skills (Domain 7), with 52% of respondents self-assessing as either 'Proficient' or 'Expert' in Domain 7. Interpersonal Skills (84%) and Evaluation Attitudes (75%) had the highest number of respondents rating themselves as Competent or higher. Respondents assessed their expertise as lowest in Theoretical Foundations (Domain 3) and Research Methods (Domain 5) with approximately 45% rating themselves as None, Novice, or Advanced Beginner.

The distribution varied across individual competencies *within* each domain, but the pattern was similar, with most of the ten highest rated competencies related to Interpersonal Skills (Domain 7) and ten lowest competencies related to Theoretical Foundations and Research Methods (Domains 3 and 5).

The ten highest rated competencies were:

- *7.2 Display empathy (61% rated as Proficient or Expert).*
- *7.1 Listen for and respect others' points of view (59%).*
- *7.9 Demonstrate the capacity to build relationships with a range of people (59%).*
- *7.4 Demonstrate effective written communication skills (57%).*
- *7.5 Demonstrate effective verbal communication skills to engage with all evaluation stakeholders (56%).*
- *7.6 Use non-verbal communication skills where relevant and appropriate (53%).*
- *7.3 Maintain an objective perspective (50%)*

- *2.14 Respect the values of others, for example, different political persuasions or social views in contrast to your own (50%).*
- *7.7 Use efficient and relevant technologies in evaluation practice (49%).*
- *2.4 Demonstrate professional credibility, discretion, and confidentiality throughout evaluation processes (49%).*

The ten lowest rated competencies were:

- *3.7 Use economic approaches to value for money assessments, for example, cost analysis, cost–benefit analysis, cost effectiveness, and social return on investment (73% rated as Advanced Beginner or less).*
- *5.12 Employ valid quantitative methods with rigour, and apply statistical analysis with awareness of the assumptions underlying the analysis (57%).*
- *3.6 Use evaluation theories, concepts, and general approaches to evaluation (53%).*
- *3.8 Undertake evaluative actions, for example, grade, rate, score, rank, apportion, compare, or attribute to establish merit or worth (53%).*
- *5.11 Design appropriate sampling methods (based on approaches such as stratified or purposive sampling, for example) to maximise learning and avoid bias (51%).*
- *4.11 Consult with stakeholders on any departure of evaluation process from cultural norms (51%).*
- *5.9 Undertake impact assessments based on the specific program or project logic and context; verify actual and perceived impacts (50%).*
- *4.12 Understand and articulate the potential limitations of the evaluation within the cultural context(s) (48%).*
- *5.5 Assess reliability and validity of data through use of data checks, control and comparison trials, triangulation of results, and cross analyses (47%).*
- *4.2 Identify and incorporate appropriate cultural protocols for interacting with the community, including incorporating cultural expertise on the evaluation team (47%).*

There were no competencies with two sided extreme distributions, that is, most people self-assessing as either Novice/Advanced Beginner or Proficient/Expert. Instead, the remaining competencies all fell within a normal distribution range with a plurality of respondents self-assessing as Competent.

Self-assessment responses, by score

To generate arithmetic means and perform statistical analyses, the authors converted respondents' individual self-assessment ratings to numerical scores as follows:

- None = 1
- Novice = 2
- Advanced beginner = 3
- Competent = 4

Table 3. Self-assessment numerical score, by Domain.

Domain	Mean score (max = 6)
Domain 1: Evaluation Activities	4.0
Domain 2: Evaluation Attitudes	4.0
Domain 3: Theoretical Foundations	3.5
Domain 4: Attention to Culture, Stakeholders, and Context	3.6
Domain 5: Research Methods	3.6
Domain 6: Evaluation Project Management	3.7
Domain 7: Interpersonal Skills	4.2

- Proficient = 5
- Expert = 6

This enabled averages (means) to be calculated for each competency rating, which could also be aggregated for each Domain (Table 3). The average scores for all the competencies can be found in Appendix B: Full results by Domain and Competency.

The highest and lowest self-assessed competencies have already been listed above but when converted to scores, the means for individual competencies ranged from 2.7 to 4.5.

Self-assessment responses, scores by demographic variable

The authors explored the relationship between self-assessment scores at the Domain level and various demographic variables of the respondents. Simple descriptive analyses were undertaken (comparison of mean Domain score) for the following variables.

1. Gender.
2. Age group.
3. Highest level of any study.
4. Formal study of evaluation.
5. Years working in evaluation.
6. Organisational context.
7. Amount of evaluation in current role.

The following charts provide results of these analyses. In cases where more than two variables are compared, the authors used column charts; line charts obscured the differences. Where Domain level differences in mean scores occurred between demographic variables, the authors performed a One-Way Analysis of Variance (ANOVA) test for statistical significance.

Male and female respondents showed no differences in average scores across the Domains. Means ranged from identical to within 0.2; no Domain had statistically significant differences ($F(1,219) = 0.37$, $p = .54$). Only two individual competencies showed an apparent score difference based on gender score. Male respondents scored themselves an average of 0.5 higher than female respondents for both 'Competency 3.7 – Use economic approaches to value for money assessments' and 'Competency – 5.12 Employ valid quantitative methods with rigour and apply statistical analysis'.

Differences in Domain scores were checked based on highest level of study and formal evaluation study. For highest level of study, there was no clear pattern or trend evident in the scores. Respondents with a PhD had the highest scores in every Domain; however, the difference was only significant in Research Methods (Domain 5) (($F (1,259) = 11.88$, $p = .001$). The average scores for all other levels of study across the Domains were very similar with no statistically significant differences ($F (1,289) = 1.22$, $p = .30$). Respondents who had studied evaluation formally and those who had not showed no apparent or significant differences in average Domain scores in any Domain ($F (1,289) = 0.38$, $p = .54$). Even at the individual competency level, there was no competency where the average score differed significantly based on whether the respondents studied evaluation formally.

Age and amount of work in evaluation yielded significant results. Comparing across age groups revealed more stark and statistically significant differences in average Domain scores ($F (3,221) = 4.65$, $p = .003$). Respondents aged 60 or above drove these results as they self-assessed higher in all Domains (Figure 4). Respondents' years working in evaluation, and percentage of evaluation in their current work role also produced statistically significant differences. For years working in evaluation ($F (3, 142) = 3.92$, $p = .01$), respondents had a median of 8 years working in evaluation (mean = 11.7 years) and there was a moderate positive correlation between years
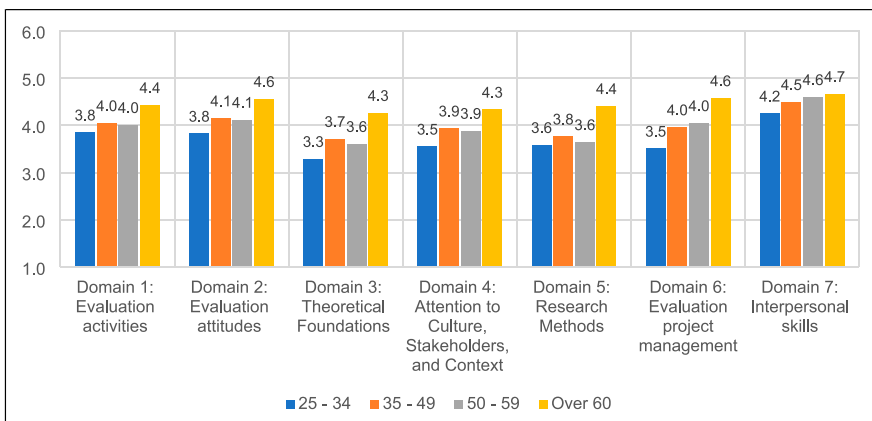
Figure 4. Average Domain score, by age group.
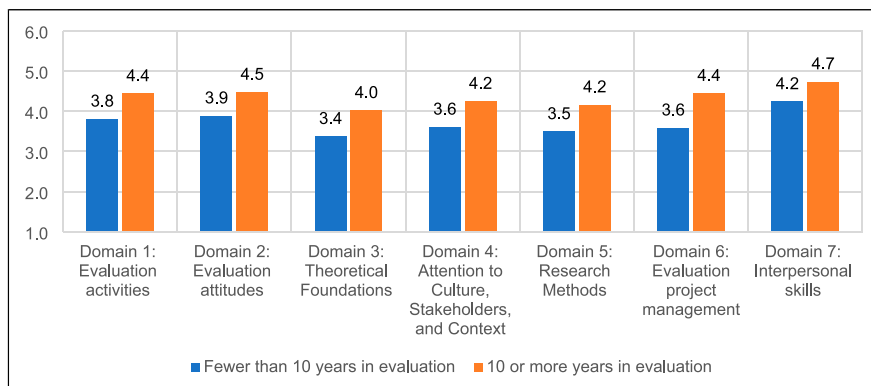
Figure 5. Average Domain score, by years working in evaluation.

working in evaluation and Domain score (r = 0.38). In every Domain, respondents who had worked in evaluation for 10 or more years had a higher average score, particularly in Research Methods (Domain 5) and Evaluation Project Management (Domain 6) (Figure 5). Where respondents' current work roles involved evaluation (F (3,142) = 3.92, *p* = .01) the difference was driven by those respondents who indicated that 50% or more of their role involved practicing evaluation. Their average scores were higher in every Domain except in Interpersonal Skills (Domain 7) where the average scores were similar. The current organisational context of respondents appeared to have little impact on the competency scores. For most Domains, those working in for-profit organisations, universities, colleges or schools, and consultancies tended to have marginally higher scores than those working in not-for-profit organisations, but these were not statistically significant (F (4, 245) = 0.36, *p* = .84).

Summary

Answers to the research questions for phase two are as follows

Research question 2.1. *Does the instrument have face and construct validity?* The findings above demonstrate that the self-assessment functioned well from a face and content validity perspective. Respondents understood the items and were able to rate their expertise, particularly after the scale was updated to include the 'none' option (see Wildschut et al., 2024, this issue for further discussion).

Research question 2.2. *What can we learn about a sample of evaluators from a competencies self-assessment?* These findings are further discussed below, but the sample was diverse enough to give us a broad range of respondents in terms of demographics, practice, age, and expertise. While the instrument will need further

analysis to begin to establish construct validity, the findings were sufficient for the AES Pathways Committee to use them to revise their workshop program, beginning in 2021.

Discussion

The authors conducted an update of the 2013 AES Evaluators' Professional Learning Competency Framework to suit online self-assessment in the LEAP. Use of the Portal self-assessment by AES members and the public generated a data set which the authors analysed to learn more about who is interested in learning evaluation and how they perceive their expertise on the competencies. The results demonstrate that most evaluation practitioners who completed the tool consider themselves competent on all seven Domains and self-assessed as competent or higher on most competencies.

Domain and competency ratings

The highest rated competencies were in Interpersonal Skills (Domain 7) and Evaluation Attitudes (Domain 2). These are not evaluation-specific, for example, '7.2 Display Empathy' or '7.4 Demonstrate effective written communication skills'. While evaluators come from a variety of educational, professional, and geographic backgrounds, as described above, interpersonal skills are a common requirement for navigating an increasingly complex global workplace (Milligan et al., 2020). As such they are found in many professional competencies lists beyond evaluation; the high ratings here may demonstrate the commonality and importance of those skills globally. However, Dewey et al. (2008) found employers agreed that evaluation job seekers over-rated their interpersonal skills and abilities: job seekers rated themselves second from the top in this area, while employers rated the seekers' interpersonal skills as third from the worst. Therefore, over-rating is possible in our study as well. To address this difficulty, Tovey et al. (2023) provide eight strategies for integrating practice with these skills into formal evaluation curricula. Their recommendations of meditation, self-assessment instruments for communication styles, reflection, and goal setting could be included in both formal and informal evaluation training. Other professions have produced validated assessment protocols for this area which could be adapted for evaluators (e.g. Duffy et al., 2004). Lovato and Shaw (2021) adapted the Team Performance Scale from medicine to serve that purpose in an evaluation subject in a Master of Public Health.

Respondents rated lowest the competencies linked to evaluation's unique knowledge base, essential for defining it as a profession (Crane, 1988; Morell & Flaherty, 1978; Worthen, 1994) and for practice. These included seven competencies related to Theoretical Foundations (Domain 3) and Research Methods (Domain 5), for example, 5.5. Assess reliability and validity of data or 3.8 Undertake evaluation actions… to establish merit or worth. One of these lower ranked competencies (3.7 Use economic approaches to value for money assessments) was originally unique internationally to the AES Framework. As it was only added recently to the AES Professional Learning

Program, a low rating is expected. For the other evaluation-specific competencies, this is likely a reflection of the available learning opportunities related to evaluation, which often focus on evaluation as it is defined and practised within a discipline or sector (Gullickson, 2020), specific evaluation tasks (e.g. creating logic models, monitoring and evaluation frameworks, and survey design), or specific approaches or techniques (e.g. developmental evaluation and most significant change) (Gullickson et al., 2019). Within the formal education space an 'evaluation-specific' credential has been characterised as including a minimum of two evaluation-specific subjects at the graduate level (LaVelle & Donaldson, 2015) as opposed to other disciplines which require dedicated coursework at undergraduate level before progressing to graduate/post-graduate study. Montrosse-Moorhead et al. (2022) found that teachers of evaluation at universities could not agree on how to differentiate skills and knowledge appropriate at master and doctoral level. LaVelle and Davies' (2021) and Pitts (2021) described content and pedagogies for introductory courses, but their research included no mention of evaluation-specific skills related to the logic of evaluation.

Attention to Culture, Stakeholders, and Context (Domain 4) held the remaining three of the ten lowest rated competencies, for example, 4.2 Identify and incorporate appropriate cultural protocols for interacting with the community. Again, this is not entirely unexpected as cultural competency has received greater focus within the AES and evaluation more broadly in the past few years. The effects of the recently developed First Nations Cultural Safety Framework (Gollan & Stacey, 2021) and trainings may not have been either captured or sufficient to improve ratings in these results. Indeed, the Framework may need updating to better reflect cultural safety knowledge and skills. Like interpersonal skills, cultural competency and cultural safety are matters of interest and investigation across disciplines and sectors. A combination of the AES Cultural Safety Framework with Australian-specific frameworks (Productivity Commission, 2020a, 2020b; https://www.lowitja.org.au/page/services/tools/evaluation-toolkit) and existing measures (e.g. Dollwet & Reichard, 2014) could be used for more targeted assessment of this competency.

While there were competencies which were left or right skewed in their distribution, that is, more novices/advanced beginners or more proficient/expert, there were no competencies with a large proportion of *both* novices/advanced beginners and proficient/expert. In other words, most respondents clustered around the 'Competent' rating. It is possible that this reflects the 'Dunning-Krueger' effect (Dunning, 2011) with less experienced/competent evaluators *over-rating* their competence and more experienced/competent evaluators *under-rating* their competence. However, the absence of any objective measurement of competency to compare against the self-assessment makes this impossible to determine.

Demographic differences

The comparison of mean Domain scores on demographic variables indicated that several had little to no relationship with self-assessed competency while others clearly made a difference. Gender, years in the workforce, work context, highest level of

education, and whether the respondent formally studied evaluation had no clear influence on mean scores. Respondents with PhDs showed higher ratings on Research Methods (Domain 5), but not elsewhere. Years working in evaluation, age, and amount of evaluation in current role were positively correlated with Domain scores. In this data set, experience, as conveyed by all three variables, has the strongest relationship with higher self-assessment. However, around 70% of LEAP respondents reported they had major gaps in their preparedness to work in evaluation. The proportion was similar across demographic variables such as age and years of evaluation experience, which suggests that individuals continue to perceive a gap in their knowledge and skills, which years of experience alone cannot resolve, and that further practical, informal or formal education and training is desired.

These findings are in accord with other studies on evaluation and evaluator competencies (e.g. Bowman (Waapalaneexkweew), 2021; Cho et al., 2022; Mason, 2020; Trevisan, 2002, 2004) – evaluators learn through experience with doing evaluation. Evaluation is at or near the top of learning taxonomies (Biggs & Collis, 1982; Krathwohl, 2002) and as such always requires the evaluator to apply knowledge from multiple domains to the current evaluation. The lack of definition about what constitutes evaluation (Gullickson, 2020), has meant that evaluation, unlike other disciplines, does not have many pathways for moving from knowing that to knowing how (Clinton & Hattie, 2021), and no clear pathways for developing expertise. This is a challenge for formal and professional educators, commissioners, and evaluation users to take up together. For evaluation to emerge as a distinct profession, it may require new pathways for evaluation learners to develop mastery in evaluation-specific and evaluation-relevant competencies, including adaptation of existing pathways from other disciplines.

Limitations

The study and report of findings has five limitations:

- In the first iteration of the LEAP there was no 'None' choice in the expertise rating scale. This may have skewed the data with people either not responding to items or choosing Novice as it was the lowest rating available.
- One of the limitations of the LEAP is that revisions of the survey required new survey links. Duplication of respondents across test versions was possible.
- The comparisons of average Domain ratings by demographic variables were subjected to statistical analyses, but this was not done at the Competency level. It is possible that the lack of statistically significant differences in Domain level ratings is masking significant differences at a Competency level based on demographic variables.
- Despite the efforts of the AES team working on the revision, some competency statements remained double barrelled. In Domain 2: Theoretical Foundations this

is particularly evident, as several of those items are unique to AES and have not been tested to see whether they should be separated out.
• Self-assessment has a variety of documented limitations, which apply to this study (see Wildschut et al., 2024, this issue).

Future research

The authors updated the AES Evaluators' Professional Learning Competency Framework throughout the development of the self-assessment. These changes and the data from the self-assessment have raised areas for future research:

• Further dissemination of the LEAP to increase the data set to discover any patterns of distribution based on disciplinary backgrounds and sectors.
• Additional updating of the competencies based on factor analysis, an analysis of value frames and paradigms, and competencies implicit in the First Nations Cultural Safety Framework.
• Identification of core competencies that differentiate evaluators from other professional areas that use interpersonal skills, research, and project management. Consideration of the duties of the evaluator (in line with 'duties of the teacher', Scriven, 1994) as an alternative to competencies to make this distinction.
• Expansion of the dataset to identify any significant differences in self-assessed ratings at a competency level, based on demographic variables, which would warrant the development of more targeted or tailored resources and training around competencies.
• Identification of whether and how professional development and learning activities can lead to more accurate self-assessment of competencies across the Domains; and acceptance that more accurate ratings may be lower or higher due to the Dunning-Kruger effect (Dunning, 2005).
• Research to determine what good evaluation looks like in different disciplines and contexts to understand its relationship to the competencies/good evaluation practice including ethics (Morris, 2015).
• Research on the effects of long-term, practical evaluation training on expertise.
• Development of robust, criterion-based assessments of competency that can accommodate the diversity of both the contexts and practitioners and the essential expertise that defines high quality evaluation practice.

Conclusion

AES and learnevaluation.org have collaborated to develop an online evaluator competency self-assessment tool based on an updated version of the 2013 AES Evaluator Professional Learning Competencies Framework. The use of the LEAP by AES and the public provided a rich dataset on the self-assessed expertise of nearly 300 evaluation practitioners. The levels of engagement with various iterations of the tool demonstrate

the interest among the evaluation community for self-assessment and identifying strengths and areas for development.

The challenges of writing competency statements for self-assessment experienced by the working group underscores the importance of measurement and the need for a measurement competency, which is currently not included in any professional association competency sets. However, the competencies can clearly be used for self-reflection and the self-assessment tool can facilitate this reflective process, providing a way for emerging and experienced evaluators alike to benchmark their assessments of competence over time.

ORCID iDs

Amy M Gullickson https://orcid.org/0000-0001-5408-9846
Delyth Lloyd https://orcid.org/0000-0001-6948-9922
Anne Stephens https://orcid.org/0000-0002-1929-4999
Lauren Wildschut https://orcid.org/0000-0002-7002-2323

Supplemental Material

Supplemental material for this article is available online.

References

AES Professional Learning Committee. (2013). *Evaluators' professional learning competency framework*. Australasian Evaluation Society. https://www.aes.asn.au/images/AES_Evaluators_Competency_Framework.pdf

Aotearoa New Zealand Evaluation Association. (2011). *Evaluator competencies*. Aotearoa New Zealand Evaluation Association. https://anzea.org.nz/assets/Key-ANZEA-Files/110801_anzea_evaluator_competencies_final.pdf.

Austin, Z. (2019). Competency and its many meanings. *Pharmacy: Journal of Pharmacy Education and Practice*, *7*(2), 37. https://doi.org/10.3390/pharmacy7020037

Australian Evaluation Society. (2021). *AES Strategic Priorities 2019-2022: Revised by the board March 2021 following a mid-term review of the strategy.* Australian Evaluation Society. https://www.aes.asn.au/images/aes_strategicPlan2019_22_finalWeb.pdf.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome).* Academic Press. https://search.ebscohost.com/login.aspx?direct=true&db=cat00006a&AN=melb.b1215135&site=eds-live

Bowman (Waapalaneexkweew), N. (2021). Praxis makes perfect? Transcending textbooks to learning evaluation experientially and in cultural contexts. *Canadian Journal of Program Evaluation*, *35*(3), 320–329. https://doi.org/10.3138/cjpe.69698

Canadian Evaluation Society. (2010). *Competencies for Canadian evaluation practice*. Canadian Evaluation Society. https://evaluationcanada.ca/files/pdf/2_competencies_cdn_evaluation_practice_2018.pdf.

Cho, M., Castleman, A. M., Umans, H., & Mwirigi, M. O. (2022). Measuring evaluator competencies: Developing and validating the evaluator competencies assessment tool. *American Journal of Evaluation*, *44*(3), 474–494. https://doi.org/10.1177/10982140211056539

Clinton, J. M., & Hattie, J. (2021). Cognitive complexity of evaluator competencies. *Evaluation and Program Planning*, *89*(November 2020), 102006. https://doi.org/10.1016/j.evalprogplan.2021.102006

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, *49*(6), 560–575. https://doi.org/10.1111/medu.12678

Coolidge, F. L., & Segal, D. L. (2010). Validity. In I. B. Weiner, & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (1st ed., pp. 1–2). Wiley. https://doi.org/10.1002/9780470479216.corpsy1019

Crane, J. A. (1988). Evaluation as scientific research. *Evaluation Review*, *12*(5), 467–482. https://doi.org/10.1177/0193841X8801200501

Dewey, J. D., Montrosse, B. E., Schröter, D. C., Sullins, C. D., & Mattox, J. R. (2008). Evaluator competencies: What's taught versus what's sought. *American Journal of Evaluation*, *29*(3), 268–287. https://doi.org/10.1177/1098214008321152

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. Wiley.

Dollwet, M., & Reichard, R. (2014). Assessing cross-cultural skills: Validation of a new measure of cross-cultural psychological capital. *International Journal of Human Resource Management*, *25*(12), 1669–1696. https://doi.org/10.1080/09585192.2013.845239

Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral vision: Expertise in real World contexts. *Organization Studies*, *26*(5), 779–792. https://doi.org/10.1177/0170840605053102

Duffy, F. D., Gordon, G. H., Whelan, G., Cole-Kelly, K., Frankel, R., Buffone, N., Lofton, S., Wallace, M., Goode, L., & Langdon, L. (2004). Assessing competence in communication and interpersonal skills: The Kalamazoo II Report. *Academic Medicine*, *79*(6), 495–507. https://doi.org/10.1097/00001888-200406000-00002

Dunning, D. (2005). *Self-insight : Roadblocks and detours on the path to knowing thyself*. Psychology Press.

Dunning, D. (2011). The Dunning–Kruger effect. *Advances in Experimental Social Psychology*, *44*, 247–296. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

European Evaluation Society. (2012). *The EES Evaluation Capabilities Framework*. European Evaluation Society. https://europeanevaluation.org/wp-content/uploads/2020/02/EES-EVALUATION-CAPABILITIES-FRAMEWORK-1.pdf.

Ghere, G., King, J. A., Stevahn, L., & Minnema, J. (2006). A professional development unit for reflecting on program evaluator competencies. *American Journal of Evaluation*, *27*(1), 108–123. https://doi.org/10.1177/1098214005284974

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2017). *Research methods in applied settings: An integrated approach to design and analysis* (3rd ed.). Taylor & Francis.

Gollan, S., & Stacey, K. (2021). *Australian Evaluation Society First Nations' Cultural Safety Framework*. Australian Evaluation Society. https://www.aes.asn.au/images/AES_FirstNations_Cultural_Framework_finalWEB_final.pdf.

Gullickson, A. M. (2020). The whole elephant: Defining evaluation. *Evaluation and Program Planning*, *79*(1), 101787. https://doi.org/10.1016/j.evalprogplan.2020.101787

Gullickson, A. M., King, J. A., LaVelle, J. M., & Clinton, J. M. (2019). The current state of evaluator education: A situation analysis and call to action. *Evaluation and Program Planning*, *75*(1), 20–30. https://doi.org/10.1016/J.EVALPROGPLAN.2019.02.012

Hodges, B. D. (2015). Sea monsters & whirlpools: Navigating between examination and reflection in medical education. *Medical Teacher*, *37*(3), 261–266. https://doi.org/10.3109/0142159X.2014.993601

Holden, R. R. (2010). Face validity. In I. B. Weiner, & W. D. Craighead (Eds.), *He Corsini Encyclopedia of Psychology* (pp. 1–2). John Wiley & Sons. https://doi.org/10.1002/9780470479216.corpsy0341

IDEAS Working Group on Evaluator Competencies. (2012). *Competencies for development evaluation evaluators, Managers, and commissioners*. International Development Evaluation Association. https://ideas-global.org/wp-content/uploads/2022/10/IDEAS-Competencies-adopted-Feb-2012.pdf

King, J. A., & Stevahn, L. (2020). Presenting the 2018 AEA evaluator competencies. *New Directions for Evaluation*, *2020*(168), 49–61. https://doi.org/10.1002/ev.20435

King, J. A., Stevahn, L., Ghere, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. *American Journal of Evaluation*, *22*(2), 229–247. https://doi.org/10.1177/109821400102200206

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

LaVelle, J. M., & Donaldson, S. I. (2015). The state of preparing evaluators. In J. W. Altschuld, & M. Engle (Eds.), *New Directions in Evaluation*, *2015*(145), 39–52. https://doi.org/10.1002/ev.20110

LaVelle, J. M., & Davies, R. (2021). Seeking consensus: Defining foundational concepts for a graduate level introductory program evaluation course. *Evaluation and Program Planning*, *88*(1), 101951. https://doi.org/10.1016/j.evalprogplan.2021.101951

Lovato, C. Y., & Shaw, G. (2021). Helping students reflect on their interpersonal skills: The Team Performance Scale (TPS). *Canadian Journal of Program Evaluation*, *35*(3), 395–402. https://doi.org/10.3138/cjpe.70376

Mason, S. (2020). Practice makes better? Testing a model for training program evaluators in situation awareness. *Evaluation and Program Planning*, *79*(1), 101788. https://doi.org/10.1016/j.evalprogplan.2020.101788

Milligan, S. K., Luo, R., Hassim, E., & Johnston, J. (2020). *Future-proofing students: What they need to know and how to assess and credential them*. Melbourne Graduate School of Education. https://www.bigpicture.org.au/files/future-proofing_students_johnson.pdf

Montrosse-Moorhead, B., Gambino, A. J., Yahn, L. M., Fan, M., & Vo, A. T. (2022). Evaluator education curriculum: Which competencies ought to be prioritized in master's and doctoral programs? *American Journal of Evaluation*, *43*(2), 269–292. https://doi.org/10.1177/10982140211020326

Morell, J. A., & Flaherty, E. W. (1978). The development of evaluation as a profession: Current status and some predictions. *Evaluation and Program Planning*, *1*(1), 11–17. https://doi.org/10.1016/0149-7189(78)90003-4

Morris, M. (2015). Research on evaluation ethics: Reflections and an agenda. In P. R. Brandon (Ed.), Research Evaluation. *New Directions for Evaluation 2015*(148), 31–42. https://doi.org/10.1002/ev.20155

Pitts, R. T. (2021). Pinpointing where to start: A reflective analysis on the introductory evaluation course. *Canadian Journal of Program Evaluation*, *35*(3), 437–449. https://doi.org/10.3138/cjpe.69697

Productivity Commission. (2020a). *A guide to evaluation under the Indigenous evaluation strategy*. Commonwealth Government. https://www.pc.gov.au/inquiries/completed/indigenous-evaluation/strategy/indigenous-evaluation-guide.pdf.

Productivity Commission. (2020a). *Indigenous evaluation strategy*. Commonwealth Government. https://www.pc.gov.au/inquiries/completed/indigenous-evaluation/strategy

Putra, W. (2015). *The validation of AES evaluators' professional learning competency framework – 2015 [Master's Capstone Report]*. The University of Melbourne.

Russ-Eft, D. F. (2008). *Evaluator competencies: Standards for the practice of evaluation in organizations*. Jossey-Bass.

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

Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, *19*(1), 1–19. https://doi.org/10.1177/109821409801900102

Smith, M. F. (1999). Should AEA begin a process for restricting membership in the profession of evaluation? *American Journal of Evaluation*, *20*(3), 521–531. https://doi.org/10.1177/109821409902000311

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation*, *26*(1), 43–59. https://doi.org/10.1177/1098214004273180

Tovey, T. L. S., Smith, L., Linnell, D. J., Wisner, D., & Coles, C. (2023). Explicitly integrating interpersonal skills in the evaluation curriculum. *New Directions for Evaluation*, *2023*(177), 23–29. https://doi.org/10.1002/ev.20541

Trevisan, M. S. (2002). Enhancing practical evaluation training through long-term evaluation projects. *American Journal of Evaluation*, *23*(1), 81–92. https://doi.org/10.1177/109821400202300110

Trevisan, M. S. (2004). Practical training in evaluation: A review of the literature. *American Journal of Evaluation*, *25*(2), 255–272. https://doi.org/10.1177/109821400402500212

Visitor Studies Association. (2008). *Evaluator competencies for professional development*. Visitor Studies Association. https://visitorstudies.org/images/resources/evaluator-competencies.pdf

Wildschut, L., Gullickson, A. M., & Mason, S. (2024). Making the Learn Evaluation Assessment Portal: Developing a reciprocal tool for learning and research. *Evaluation Journal of Australasia*, *24*(4), 243–271. https://doi.org/10.1177/1035719X241293805

Worthen, B. R. (1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? *New Directions for Program Evaluation*, *1994*(62), 3–15. https://doi.org/10.1002/ev.1672

Worthen, B. R., & Sanders, J. R. (1991). The changing face of educational evaluation. *Theory into Practice*, *30*(1), 3–12. https://doi.org/10.1080/00405849109543470

Yarbrough, D. B., Caruthers, F. A., Shulha, L. M., & Hopson, R. K. (2010). *The program evaluation standards: A guide for evaluators and evaluation users*. Sage.