Professionalism in Program Evaluators: A Comparison of American and Canadian Evaluators

Sandra Ayoo
University of North Carolina, North Carolina, United States

Abstract: The professionalization of evaluation means different things to different people, and as a result, the field lacks a clear understanding of how to empirically assess evaluator professionalism. This exploratory study used a sociological model to study the behaviours of practicing evaluators in Canada and the United States using five concepts: expertise, ethical disposition, professional autonomy, innovation and research, and credentialing. Results from 27 in-depth interviews and a survey of 456 respondents demonstrated that, depending on their demographic characteristics, perceptions of professionalism differed between evaluators in the two countries. The study determined that theory-based sociological models can provide insights into the professionalization of program evaluation.

Keywords: credentialing, ethics, expertise, innovation and research, professional autonomy, professionalism, professionalization

The international evaluation community welcomed the historic declaration by the United Nations (UN) that 2015 was the international year of evaluation. This historic declaration was important for two reasons. First, the UN profiled program

Corresponding Author: Sandra Ayoo, University of North Carolina Greensboro, School of Education Building, Room 246, 1300 Spring Garden St. Greensboro, North Carolina, 27412, United States; sjayoo@uncg.edu

© 2022 Canadian Journal of Program Evaluation / La Revue canadienne d'évaluation de programme
doi: 10.3138/cjpe.71300
This advance access version may differ slightly from the final published version
evaluation globally as an indispensable discipline needed for the successful implementation of the Sustainable Development Goals (SDGs) launched in 2015. Member states highly criticized the Millennium Development Goals (MDGs) that ended in 2015 for not paying sufficient attention to evaluative evidence on achievement of the goals and knowledge gained from implementation challenges (IOCE, n.d.). Second, the declaration led to reflection by evaluators worldwide, leading to a movement to professionalize program evaluation (Ayoo, 2020).

The International Organization for Cooperation in Evaluation (IOCE) and EvalPartners are leading the global movement to professionalize evaluation, using “professionalism” as their guiding framework. The IOCE sees professionalism as the outcome of a collective endeavour carried out by an occupational group to improve the relevance, quality, and delivery of its expert services in the public interest (IOCE, n.d.), but in general it is not clear what the term professionalism in evaluation means (Schwandt, 2017) or how to empirically assess it (Ayoo, 2020). It seems reasonable to claim that professionalism is the outcome of professionalization, the collective endeavour referred to by the IOCE. Stated differently, professionalization is the process by which an occupation becomes a profession (Schwandt, 2017), and professionalism is the outcome of that process.

I examined evaluator professionalism in the United States and Canada using a sociological model of professionalism advanced in the study of professions. The purpose of this study was to investigate both the extent to which and the ways in which program evaluation is becoming a profession in Canada and the United States by studying the self-reported behaviours of evaluators on five professionalism scales. Two research questions were examined:

**Research Question 1**: How do program evaluators in North America describe professionalism in their everyday practice of evaluation?

**Research Question 2**: To what extent and in what ways are program evaluators in Canada and the United States demonstrating professionalism in their practice of evaluation?

I used the theoretical assertions in the definitions of professionalism to hypothesize that evaluators’ self-reported professionalism would differ by their years of experience, credential status, and workplace. To ground this investigation, it is important to provide an overview of the history of professionalization in the United States and the history of the Professional Designation Program (PDP) in Canada.

**BRIEF HISTORY OF PROFESSIONALIZATION OF EVALUATION IN NORTH AMERICA**

Sociological studies of professions are replete with accounts of how occupations become professions (see Davidson-Shivers & Barrington, 2004), and the theory of the occupation-profession continuum (Pavalko, 1988) is a useful heuristic device for understanding the history of the professionalization of program evaluation in the United States and Canada. The most common traits of professions are
that they are a full-time occupation; have skills based on theoretical knowledge, training, and education; have tests to demonstrate competence; adhere to a code of conduct; provide a public good; and are organized through the formation of associations (see Davidson-Shivers & Barrington, 2004). Although these traits are not considered linear in their progression, Pavalko (1988) argues that occupations at the higher end of the continuum have more of the defining traits of professions.

Applying these concepts to evaluation helps to understand how program evaluation is evolving from an occupation to a profession in the two countries. House (1990) writes that program evaluation has transformed from a small, sideline activity conducted by part-time academics into a “professionalized minor industry, replete with its own journals, awards, conventions, organizations, and standards” (p. 24) and that evaluation has its own entry in the Dictionary of Modern Thought. This behavioural practice has evolved over the last half a century into a field of professional practice, with more than 138 professional associations of evaluators and voluntary organizations for professional evaluation (VOPEs) of various capacities existing in more than 110 countries (IOCE, n.d.). Some scholars (e.g., Sou, 2008) suggest that program evaluation originated from the practice of educational assessments of programs and student performances traced to the “Payment-by-Results” (PBR) schemes originally developed in Italy in 1444. Madaus et al. (2012) provide a detailed account of the history of program evaluation in the United States framed in eight historical timelines or stages.

As shown in Figure 1, before 1965, evaluation in the United States meant testing or curriculum evaluation, mainly by measurement experts, but it then became a multidisciplinary field with separate traditions (House, 1990). Two significant events in the history of evaluation in the United States were the enactment of the 1958 National Defense Education Act (NDEA) following the national crisis sparked by Russia’s launch of Sputnik I and the passage of the 1965 Elementary and Secondary Education Act (ESEA), which stimulated large-scale evaluation studies of federal government grants (Russ-Eft & Preskill, 2009). Since 1965, evaluation has become part and parcel of most federal grants (King, 2003; Russ-Eft & Preskill, 2009), positioning evaluation as an emerging profession in the United States.

Similar to the situation in the United States, Gauthier et al. (in press) argue that program evaluation in Canada is driven by the federal government sector in its “evolution, structure, and practice” (p. 4). These authors observed that the Canadian federal government has adopted at least five evaluation policies over four decades; first introduced in 1977, “the Policy has changed four times over the past 40 years—in 1991, 2001, 2009, and 2016” (p. 5). As can be seen in Figure 1, it is interesting to note that no major events happened in the professionalization process in either country for nearly a decade in the 1970s, almost until the formation of the professional associations in the 1980s. Although the Canadian Evaluation Society (CES) is the oldest professional association in North America, as seen in Figure 1, in fact, the Evaluation Research Society (ERS) and the Evaluation Network in the United States predated it. They merged to form the American Evaluation Association (AEA) in 1986. The formation of professional associations

doi: 10.3138/cjpe.71300 e71300, CJPE © 2022

This advance access version may differ slightly from the final published version
triggered other important events in the professionalization of program evaluation in both countries. These included the adoption of professional guidelines and standards of practice, like the Program Evaluation Standards and the ethical guidelines for evaluators in the early to mid-1990s. In 2001, the development of the Essential Competencies for Program Evaluators (ECPE), independent of CES and AEA (King & Stevahn, 2015; Stevahn et al., 2005), was a significant event in the professionalization process of program evaluation in North America and around the globe.

The PDP was preceded by the CES adoption of the evaluator competencies in 2008 after fact-finding and member-wide consultations on evaluator credentialing by a consortium of prominent and leading members of the Canadian evaluation community in 2006 (Cousins et al., 2009). Importantly, the CES Board approved the launch of the historic Credentialed Evaluator (CE) Professional Designation Program (PDP) in 2010 (Kuji-Shikatani et al., 2015), a decade after the publication of the ECPE. In comparison to Canada where significant steps were taken, including establishing the PDP, the 2000s was a quiet decade for the AEA regarding professionalization, aside from the adoption of the Statement of Cultural Competence in 2011. The AEA Board adopted evaluator competencies in 2018, 10 years after CES and 17 years after the ECPE, as shown in Figure 1. In fact, CES revised its evaluator competencies in 2018 after the evaluation of the PDP in 2015. A major event in the United States was the federal government’s 2018 enactment of the Foundations for Evidence-Based Policymaking Act (2018), which is geared to formalizing the role of monitoring and evaluation in decision making. The AEA made significant contributions to the Evidence-Based Act of 2018 (AEA, n.d.).

THEORETICAL FRAMEWORK FOR PROFESSIONALISM

As noted previously, professionalization is the process by which an occupation becomes a profession (Schwandt, 2017), and this section presents professionalism as the outcome of that process. Figure 2 presents the sociological constructs of professionalism to help frame the definition of professional evaluator in the United States and Canada. Ronald M. Pavalko (1988), an American sociologist renowned for his work on the professions, observed nearly five decades ago that the use of the term professional connotes experience, competence or expertise, licensure, and full-time performance of an activity or occupation like program evaluation for pay. In turn, Freidson (1994) argued that what distinguishes professions from occupations is the degree of professionalism, which can be empirically established by studying the behaviours of professionals. He theorized that professionalism comprises service to others, expertise, control of work or autonomy, and innovation and research by members of a profession to improve it. Despite these early attempts at defining professionalism, more recent work notes a lack of agreement on the meaning or definition of professionalism or how to measure it (Irby & Hamstra, 2016).

This lack of agreement has led to varied efforts to develop professionalism self-assessment tools within diverse fields, such as medicine (Brody & Doukas,
Figure 1. A brief history of the development of program evaluation in the US and Canada since the 1960s

Note. The Canadian timeline shows a brief historical pathway of professional development between the 1980s and 2018 (Kishchuk et al., 2019)
Ayoo 2014), pharmacy (Kelley et al., 2011), and nursing (Lombarts et al., 2014). Given this disciplinary focus, the items in these instruments do not speak to the practice of program evaluation. This research extends previous work by Modarresi (1998), which assessed the structural and attitudinal dimensions of professionalism in sampled members of AEA. In the present study, I assessed the behavioural attributes of professionalism in evaluation using five concepts from the sociological literature.

Following Figure 2, in this investigation, professional autonomy refers to the freedom professionals have from clients and the employing organization to perform their duties (Forsyth & Danisiewicz, 1985). In theory, evaluators must be free to frame the evaluation without undue pressure or interference from the commissioners or funders of the evaluation. Expertise relates to technical competence to perform program evaluation work, which can be obtained through specialized education at the tertiary level followed by substantial exposure to expert practice (Freidson, 1994; Picciotto, 2011). Research and innovation relate to experimentation done by members of associations to find new practical solutions for practice (Freidson, 1994). Ethical dispositions are measures set in place to protect the public, especially codes of ethics, disciplinary committees, and peer or collegial review (Freidson, 2001; Picciotto, 2011). Finally, credentialing refers to tested performance (Freidson, 1994; Picciotto, 2011) or the process that “specifies that a person has the requisite knowledge, skills, and practical experiences to be deemed worthy of a credential or designation of such status” (Altschuld & Engle, 2015, p. 6). Status and prestige were beyond the scope of this study due to time

![Figure 2. The professionalism constructs](image-url)
Considering the dearth of theoretical and methodological scholarship on evaluator professionalism in the literature, I used an exploratory sequential mixed methods design (Creswell & Plano Clark, 2018) to study the professionalism behaviours of practicing evaluators. As shown in Figure 3, the first phase of the study was the development of a theoretical framework noted in the previous section. The five theoretical constructs are expertise, innovation and research, ethics, professional autonomy, and credentialing. Second, the qualitative data collection and analyses were prioritized due to the newness of this research in the field of evaluation. I interviewed a purposive sample of 27 information-rich practicing evaluators in Minnesota and five provinces in Canada to co-create the meanings and to operationalize the theoretical constructs. Minnesota was selected given the longstanding development of evaluators in the state. The interviewees worked in government and non-profit organizations, independent consulting, and academic institutions, including past presidents of the AEA and CES.

Third, the practitioners’ operational definitions of the constructs were used to develop a new Professionalism Assessment Tool for Program Evaluators (PATPE). I combined the items from the emic views of participants in the qualitative study with those from existing professionalism scales to build the instrument. The items in the initial draft instrument were reviewed for face validity by senior Evaluation Studies faculty at a large American institution. In addition, the content validity was reviewed by a panel of eight subject-matter experts drawn from AEA, CES, and the European Evaluation Society. I field-tested the final instrument in the fourth phase using a wider sample of 1,000 randomly selected practicing evaluators pulled from the AEA membership and 573 anglophone CES members in 2017. These individuals were contacted via email sent from Qualtrics, and three reminder emails were sent after intervals of one week. Participants were asked to respond to statements that queried their feelings regarding professionalism on a four-point Likert-type scale with a Not Applicable option. The scales ranged from Strongly Disagree to Strongly Agree as follows: Strongly Disagree = 1, Disagree = 2, Agree = 3, and Strongly Agree = 4. The Not Applicable = 5 responses were extremely few and were therefore excluded from multivariate analyses and interpretation.

Out of the random sample (N = 1,573) provided by AEA and CES, nine emails bounced back, one person opted out, and 90 international members of AEA were excluded, resulting in a sampling frame of 1,473. Overall, 366 individuals completed the survey, 248 Americans and 118 Canadians (a 24.8% response rate). This response rate was deemed acceptable, based on Dillman et al. (2014). The technical details of the process of developing the measurement instrument...
and the psychometric analyses are presented in Ayoo (2020). Briefly, the exploratory factor analyses confirmed a presence of five latent factors measuring professionalism. These factors were validated using confirmatory factor analyses and structural equation modeling (SEM). The SEM modifications produced a model with excellent statistical fit indices and satisfactory reliability (α = 0.90, Omega = 0.95).

I analyzed qualitative data using a mix of inductive and deductive approaches (Corbin & Strauss, 2008), using MAXQDA 2019 ((VERBI Software, 2019) to interpret meanings. The quantitative data were analyzed using IBM SPSS Statistics for Windows, Version 24, to conduct Multivariate Analysis of Variance (MANOVA) and post-hoc comparisons to determine whether there were any differences between hypothesized groups on professionalism scales. Possible relationships between the scales were tested using Pearson’s correlation.

Interpretation was the final phase, in which I linked together the qualitative and quantitative results to create a more holistic understanding of the findings using joint display matrices to integrate data (Fetters & Molina-Azorin, 2017). The interpretation of the findings was based on the theoretical assertions that I had hypothesized, except for the assertion that professionalism connotes licensure. This was not assessed, since program evaluators are not licensed to practice in either the United States or Canada. Instead, I assessed whether self-reported professionalism differed by credentialed evaluator status, since Canada rolled out voluntary credentialing in 2010, eight years prior to data collection.

**KEY FINDINGS**

**Survey demographics**

The respondents represented evaluators from various employment backgrounds: internal evaluators working with governments, foundations, and non-profits; external evaluators working in the private sector or self-employed; faculty in academia; and commissioners or managers. Overall, more than half of the participants identified as Americans (64%) and about one-third as Canadians (30%); nearly four-fifths of the American (77%) and Canadian (78%) respondents identified as white, and a large proportion of these participants (62%) were female (see details in Appendix). Nearly two-thirds identified as proficient/skilled evaluators, compared to one-fifth who identified as expert/master evaluators, and 90% had postgraduate education. Participants ranged in age from 24 to 65 years old.

**Emic views of participants on professionalism**

The first research question explored how program evaluators described the sociological attributes of professionalism in their everyday practice. The qualitative study participants defined expertise as technical competence or proficiency in conducting program evaluation. These participants reported developing expertise
Figure 3. Procedural diagram of an exploratory sequential mixed methods study of professionalism in program evaluation

**RQ1.** How do program evaluators describe professionalism in their everyday practice of evaluation?

- **Theory** Constructs of professionalism
  - Literature review
  - Conceptual framework
  - Existing instruments

- **Qualitative data collection**
  - Interview 27 evaluators
  - Item generation from key quotes
  - Product: Pool of 168 items

- **Build new measurement instrument**
  - Face validity by faculty committee
  - Content validity by panel of experts
  - Interview 3 experts
  - Product: 59 items

- **Quantitative data collection and analyses**
  - Item reduction
  - Factor analyses
  - Multiple analysis of variance
  - Product: 42 items

**RQ2.** To what extent and in what ways do program evaluators demonstrate professionalism in their practice of evaluation?

- **Interpretation**

*Note.* The figure includes the research questions for each phase of the study.
through ongoing “self-assessment” of their strengths, engaging in experiential learning or “learning-by-doing,” participating in “on-the-job training,” and learning from thought leaders in the field. Nearly every evaluator interviewed cited improving their skills in evaluation through “reading on their own” and facilitating or attending “professional development” activities, usually through professional associations.

The interview participants defined innovation and research in terms of methodological and theoretical innovations to develop practical solutions to advance the discipline. Nevertheless, several evaluators in the sample noted that conducting research on evaluation to advance the field is a large part of the job requirement for academic evaluators, and they therefore argued that practitioners are less suited for research or lacked the funds for it. The key informant interviews pointed to the lack of funding for empirical research on evaluation (along the lines of Michael Scriven’s Faster Forward Fund, for example) and called on professional associations to support innovation and research through partnerships with the private sector or major foundations.

The interview participants defined professional autonomy as the freedom to provide technical oversight and being able to turn down jobs or end participation in an evaluation when negotiations with the client or employer were not fruitful. Several evaluators noted that professional autonomy is a matter of balance and compromise, since evaluators need to collaborate and negotiate with stakeholders throughout the process. Nevertheless, participants observed that power dynamics affected evaluators’ autonomy, especially for novice and younger evaluators. Expert evaluators reported using their professional association guidelines, such as AEA’s Guiding Principles and the Program Evaluation Standards, as shields from undue interferences. However, participants lamented that the AEA’s and CES’s ethical statements were good on paper but lacked enforcement by the professional associations.

Study participants defined ethics in terms of operational attributes of virtue, practical wisdom, and evaluator identity. They noted that practicing evaluation requires integrity, being culturally competent, accountable, not doing harm, and doing the right thing regardless of the consequences. Participants reported taking responsibility in the provision of information to clients to facilitate their decision making and bringing their best professional judgment within the standard of evaluation practice.

Participants had different feelings about credentialing. While 10 evaluators credentialed by the CES PDP noted that credentialing encouraged them to reflect and to self-assess their skill sets, with the goal of becoming a competent evaluator, four evaluation managers believed that credentialing would assure them that those who engage in evaluation conform to certain professional standards. Interestingly, some participants with doctoral degrees said credentialing would show their clients and employers that they had gone through their professional association’s criteria to practice evaluation, thus embracing credentialing as a source of professional identity.
**Professionalism by credentialed status**

The second research question investigated the extent to which and in what ways program evaluators in Canada and the United States are demonstrating professionalism in their practice of evaluation. The MANOVA showed that evaluators’ credential status was statistically significantly related to their perceived professionalism ($F(5, 323) = 11.982, p < 0.0005; \text{Pillai's trace } \eta^2 = 0.156$). Therefore, the hypothesis that credentialed evaluators (CEs) were more likely than non-credentialed evaluators (non-CEs) to define themselves as evaluation professionals was confirmed. It is not surprising that Canadian evaluators in the study reportedly felt that credentialing was more important on average compared to American evaluators. However, the study participants were concerned that voluntary credentialing lacked weight without endorsement by key stakeholders or if it was not required to secure a job or contract. For example, some participants noted in the interviews that

> With voluntary credentialing, if you can get organizations like the World Bank and the federal government and state governments to give extra points for credentials, voluntary credentials, that's how it would gain some momentum.

(Interview respondent, July 26, 2017)

> If suddenly I was going to lose contracts because I couldn't prove that our staff was credentialed, then we would have to do it, but otherwise, I wouldn't do it.


There were no statistically significant differences in innovation and research or ethics by credentialed status.

**Professionalism by evaluators' workplace**

Results showed a significant relationship between an evaluator's workplace and their perceived professionalism ($F(15, 922.428) = 5.111, p < 0.0005; \text{Wilks' } \Lambda = 0.802; \text{partial } \eta^2 = 0.071$). Therefore, the hypothesis that evaluator professionalism behaviours and their attitudes to credentialing had a relationship with their workplaces was confirmed. Table 1 shows statistically significant relationships between evaluators’ workplace and perceived innovation and research, professional autonomy, and overall professionalism. The quantitative study showed moderate positive correlations between innovation and professional autonomy, $r(404) = 0.358, p = 0.0005$, and between expertise and innovation, $r(430) = 0.584, p = 0.0005$. The correlations suggested that evaluators who engaged in innovation and research reported strong professional autonomy and expertise traits. It is not surprising to see from the post-hoc analyses that academic evaluators reported higher perceptions of innovation and research on average than other evaluator groups. Furthermore, external and academic evaluators reported higher perceptions of professional autonomy from clients and employers on average than internal evaluators and evaluation managers.

*doi: 10.3138/cjpe.71300 e71300, CJPE © 2022*
These findings confirmed those from the qualitative study: Academic evaluators with access to money and links to universities were more likely than non-academic evaluation practitioners to innovate novel approaches to promote the development of the profession. In fact, 77.8% ($n = 63$) of academic evaluators reported participating in formal research studies on evaluation, and 63% ($n = 51$) reported purposefully generating new evaluation knowledge to benefit the field. It is not surprising to see that internal evaluators reported the lowest perception of professional autonomy when compared with other evaluator groups. Post-hoc tests also revealed that academic evaluators were more likely than non-academic evaluators to identify as professional evaluators.

**Professionalism by evaluators’ years of experience**

Results showed a significant relationship between an evaluator’s experience in years and their perceived professionalism ($F (15, 1125) = 6.738, p < 0.001$; Pillai’s $\eta^2 = 0.247$; partial $\eta^2 = 0.115$). Therefore, the hypothesis that evaluators’ professionalism behaviours and their beliefs in credentialing were related to their experience in years of evaluation practice was confirmed. Table 2 shows statistically significant relationships between evaluators’ years of experience with innovation and research, professional autonomy, credentialing, and expertise. While the post-hoc tests showed that novice evaluators with 10 years or less of evaluation experience were, on average, more likely to believe in credentialing as a good thing than more experienced evaluators, evaluators with more than 10 years’ experience were, on average, more likely to report significantly higher perceptions of innovation and research, expertise or technical competence, professional autonomy, and overall professionalism, suggesting that these traits could be gained with years of evaluation experience. However, there were no statistically significant differences in evaluator years on ethical dispositions.

**DISCUSSION AND IMPLICATIONS FOR FUTURE RESEARCH**

This exploratory study investigated how evaluation is becoming a profession in Canada and the United States by studying the self-reported behaviours of practicing evaluators on five traits of professionalism, extending previous work on evaluator professionalism by Modarresi (1998). The description of key sociological professionalism traits (Freidson, 1994, 2001; Pavalko, 1988) and self-assessment by the participants can begin to suggest who a professional evaluator is and the level of maturity of evaluation as a field of professional practice. Because this framework encompasses frequently mentioned aspects of professionalism, the findings provide new insights about the current discourse among evaluators on the professionalization of the field.

Results from 27 interviews and a survey of 456 respondents demonstrated that reported professionalism differed by their demographic characteristics. Interestingly, these results suggest that views of being a professional evaluator depend on years of experience, credential status, and workplace. For example, while...
Table 1. Univariate mean differences by evaluator workplace (at $\alpha < 0.05$ level)

|                  | Error df | External | | Internal | | Academics | | Managers | | F | | Partial eta squared | | p | | Observed power |
|------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Innovation       |          | $M$ 2.75 | $SD$ 0.53 | $M$ 2.57 | $SD$ 0.61 | $M$ 2.99 | $SD$ 0.59 | $M$ 2.58 | $SD$ 0.64 | 8.826 | 0.073 | 0.001*** | 0.976 |
| Ethics           |          | $M$ 3.46 | $SD$ 0.45 | $M$ 3.32 | $SD$ 0.54 | $M$ 3.38 | $SD$ 0.60 | $M$ 3.33 | $SD$ 0.66 | 1.462 | 0.013 | 0.022 | 0.182 |
| Autonomy         |          | $M$ 2.78 | $SD$ 0.68 | $M$ 2.17 | $SD$ 0.83 | $M$ 2.67 | $SD$ 0.72 | $M$ 2.42 | $SD$ 1.01 | 13.350 | 0.106 | 0.001*** | 0.999 |
| Credentialing    |          | $M$ 2.60 | $SD$ 0.72 | $M$ 2.82 | $SD$ 0.64 | $M$ 2.83 | $SD$ 0.78 | $M$ 2.95 | $SD$ 0.59 | 3.471 | 0.030 | 0.016 | 0.556 |
| Expertise        |          | $M$ 3.16 | $SD$ 0.52 | $M$ 3.05 | $SD$ 0.57 | $M$ 3.10 | $SD$ 0.55 | $M$ 3.09 | $SD$ 0.64 | 0.730 | 0.006 | 0.535 | 0.073 |
| Professionalism  |          | $M$ 2.95 | $SD$ 0.37 | $M$ 2.82 | $SD$ 0.40 | $M$ 3.01 | $SD$ 0.38 | $M$ 2.89 | $SD$ 0.46 | 4.193 | 0.036 | 0.006*** | 0.671 |

Note: *$p < 0.05$ **$p < 0.01$ ***$p < 0.001$. The effect size ($d$) was computed by MANOVA. Partial eta squared ($\eta^2$) 0.000 to 0.003 = no effect, 0.010 to 0.039 = small effect, 0.060 to 0.110 = medium or intermediate effect, 0.140 to 0.2 = large effect, and 0.039 to 0.2 = desired effects (Cohen, 1988).

Table 2. Univariate mean differences by years of evaluation experience (at $\alpha < 0.05$ level)

|                  | df | 10 years or less | | 11–20 years | | 21–30 years | | 31 years or more | | F | | d (Eta squared) | | p | | Observed power |
|------------------|----|------------------|----------|------------------|----------|------------------|----------|------------------|----------|----------|----------|----------|----------|
| Innovation       |    | $M$ 2.59 | $SD$ 0.60 | $M$ 2.80 | $SD$ 0.51 | $M$ 2.90 | $SD$ 0.54 | $M$ 2.99 | $SD$ 0.47 | 8.534 | 0.064 | 0.001*** | 0.971 |
| Ethics           |    | $M$ 3.31 | $SD$ 0.54 | $M$ 3.46 | $SD$ 0.45 | $M$ 3.47 | $SD$ 0.48 | $M$ 3.34 | $SD$ 0.56 | 2.809 | 0.022 | 0.039 | 0.436 |
| Autonomy         |    | $M$ 2.22 | $SD$ 0.78 | $M$ 2.76 | $SD$ 0.77 | $M$ 2.71 | $SD$ 0.80 | $M$ 2.97 | $SD$ 0.62 | 17.547 | 0.123 | 0.010*** | 1.000 |
| Credentialing    |    | $M$ 2.85 | $SD$ 0.63 | $M$ 2.70 | $SD$ 0.74 | $M$ 2.51 | $SD$ 0.90 | $M$ 2.43 | $SD$ 0.83 | 5.682 | 0.043 | 0.001*** | 0.842 |
| Expertise        |    | $M$ 2.93 | $SD$ 0.54 | $M$ 3.20 | $SD$ 0.46 | $M$ 3.30 | $SD$ 0.50 | $M$ 3.25 | $SD$ 0.46 | 12.310 | 0.089 | 0.001*** | 0.998 |
| Professionalism  |    | $M$ 2.81 | $SD$ 0.37 | $M$ 2.99 | $SD$ 0.36 | $M$ 2.99 | $SD$ 0.41 | $M$ 2.99 | $SD$ 0.40 | 6.852 | 0.052 | 0.001*** | 0.977 |

Note: *$p < 0.05$ **$p < 0.01$ ***$p < 0.001$. The effect size ($d$) was computed by MANOVA. Partial eta squared ($\eta^2$) 0.000 to 0.003 = no effect, 0.010 to 0.039 = small effect, 0.060 to 0.110 = medium or intermediate effect, 0.140 to 0.2 = large effect, and 0.039 to 0.2 = desired effects (Cohen, 1988).
academic, long-serving, and CEs identified themselves as professional evaluators, younger and novice evaluators viewed credentialing as a positive goal for their future.

Importantly, the finding about CEs is consistent with studies from both Gauthier et al. (2015) and Fierro et al. (2016). One possible explanation for the difference in the CEs’ professional identity could be that going through the credentialing process and maintaining an active CE designation triggered professional identity formation (PIF)—a process explained in the medical education literature by the conceptual change theory (Kay et al., 2019). These scholars argue that changes in personal conceptualizations are initiated when different experiences introduce cognitive disequilibrium. It could be that the process of putting together a portfolio, conscious and deliberate reflection on evaluator competencies, and responding to feedback from experienced and well-regarded reviewers challenged the CE applicant’s concept of self and triggered a new conceptualization of self as a professional as they received their CE designation. Future research is needed to assess how going through the credentialing process may trigger professional identity formation.

While the findings that newer and novice evaluators endorsed credentialing more so than the long-serving evaluators is contrary to Lawson et al.’s (2020) study, which found that the older, more experienced, and highly educated full-time evaluators were most likely to be CEs in Canada, it appears the grandfathering effect might have affected their findings. Nevertheless, the findings from the current research endorses Lawson et al.’s (2020) observation that the younger demographics would benefit the most from the CE designations than those already well into their careers and highly educated. This implies that novice and evaluation students could learn professionalism traits if intentionally integrated into the training programs like practicums and professional development courses designed by professional associations.

With regard to expertise, the findings from the current investigation challenge the notion that evaluators gain competence in the mastery of approaches through long-term specialized tertiary education (Picciotto, 2011); rather, they appear to gain competence through deliberate practice or informal learning, consistent with previous work by other scholars that most evaluators gained mastery of their trade through informal training (Johnson, 2018; LaVelle & Donaldson, 2015). This finding is important for evaluator education since it appears that the field of program evaluation does not have a uniform core body of knowledge for pre-service training. In fact, the survey results from this research suggest that evaluator competencies should provide a framework for teaching evaluation, similar to findings from Galport and Azzam’s (2017) study. The use of competency-based training for evaluators is a prime area for future research, especially regarding how different competencies are taught and learned.

This research has shown the importance of using a professionalism framework to understand the dimensions of evaluator professionalization—a framework that might support professional associations in designing courses, training,
and other interventions to support their members in attaining higher levels of professionalism. As we envision program evaluation as a field of professional practice, professional associations must plan for ways to position it as a public good by holding practitioners accountable for their actions to improve evaluator professionalism.

NOTE
1. Information-rich cases are “cases from which one can learn a great deal about the focus of inquiry and which therefore are worthy of in-depth study” (Patton, 2015, p. 308).

REFERENCES

doi: 10.3138/cjpe.71300  e71300, CJPE © 2022
This advance access version may differ slightly from the final published version


**AUTHOR INFORMATION**

**Sandra Ayoo** is an evaluation teacher, researcher, and practitioner who is passionate about promoting program evaluation as a field of professional practice. Her scholarship focuses on studying the behaviours of evaluators to assess their level of professionalism and conducting meta-evaluations to assess the quality of evaluations. Her goal is to enhance the pre-service training of evaluators to improve the quality of evaluation service and to promote equity, diversity, and inclusion in evaluation practice.
### APPENDIX: SURVEY RESPONDENTS’ DEMOGRAPHIC CHARACTERISTICS

<table>
<thead>
<tr>
<th>Demographic variable</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>103</td>
<td>22.6</td>
</tr>
<tr>
<td>Female</td>
<td>282</td>
<td>61.8</td>
</tr>
<tr>
<td>Not specified</td>
<td>71</td>
<td>15.6</td>
</tr>
<tr>
<td>Nationality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American</td>
<td>248</td>
<td>54.4</td>
</tr>
<tr>
<td>Canadian</td>
<td>118</td>
<td>25.9</td>
</tr>
<tr>
<td>Other (Please specify)</td>
<td>90</td>
<td>19.7</td>
</tr>
<tr>
<td>CES Credentialed Evaluator designation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>44</td>
<td>9.6</td>
</tr>
<tr>
<td>No</td>
<td>298</td>
<td>65.4</td>
</tr>
<tr>
<td>Not specified</td>
<td>114</td>
<td>25.0</td>
</tr>
<tr>
<td>Primary professional identity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manager</td>
<td>35</td>
<td>9.8</td>
</tr>
<tr>
<td>Faculty/instructor</td>
<td>81</td>
<td>22.8</td>
</tr>
<tr>
<td>External</td>
<td>132</td>
<td>37.1</td>
</tr>
<tr>
<td>Internal</td>
<td>108</td>
<td>30.3</td>
</tr>
<tr>
<td>Professional association membership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canadian Evaluation Society</td>
<td>98</td>
<td>26.7</td>
</tr>
<tr>
<td>American Evaluation Association</td>
<td>244</td>
<td>66.5</td>
</tr>
<tr>
<td>Canadian Evaluation Society and American Evaluation Association</td>
<td>25</td>
<td>6.8</td>
</tr>
<tr>
<td>Years of experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 years or less</td>
<td>187</td>
<td>46.9</td>
</tr>
<tr>
<td>11–20 years</td>
<td>104</td>
<td>26.1</td>
</tr>
<tr>
<td>21–30 years</td>
<td>68</td>
<td>17.1</td>
</tr>
<tr>
<td>31 years or more</td>
<td>36</td>
<td>9.1</td>
</tr>
<tr>
<td>Not applicable</td>
<td>3</td>
<td>0.8</td>
</tr>
</tbody>
</table>