Contents lists available at SciVerse ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Further considerations of evaluation competencies in Taiwan

CrossMark

Yi-Fang Lee [a,1], James W. Altschuld [b,*], Lung-Sheng Steven Lee [a,2]

[a] National Taiwan Normal University, 162 HePing East Road Section 1, Taipei 10610, Taiwan
[b] The Ohio State University, 3253 Newgate Court, Dublin, OH 43017, United States

A B S T R A C T

A list of evaluator competencies (Stevahn, King, Ghere, & Minnema, 2005) was adapted to fit the Taiwanese context by Lee, Altschuld, & Lee (2012). It was studied as to how it generalized to a large sample in Taiwan. Likert and Fuzzy surveys with needs assessment formats (importance and competence) were mailed via random assignment to two groups of participants. The questions for the study were: do the modified competencies relate country-wide to Taiwan, did the investigation uncover training needs for evaluators, and were there convergent rating patterns across the two forms of the instrument? The results supported a fit of the modified competencies to the context and convergent validity was observed but strong competency needs were not apparent. Reasons for the findings and implications for future research are discussed.

1. Introduction

The field of educational program evaluation (EPE) is progressing toward a profession around the world. This is true for the Asia-Pacific region but at the same time concerns have arisen there about the quality of evaluations, shortages of experienced, trained evaluators, and identification of the skills or competencies for qualified persons (Garden, 2010; Hay, 2010; Hung, Altschuld, & Lee, 2012; Kumar, 2010; Lee, Altschuld, & Hung, 2008).

Accordingly in 2011, Lee et al. (2012) conducted research in the region related to the topic. It was based on the work of Stevahn et al. (King, Stevahn, Ghere, & Minnema, 2001; Stevahn, King, Ghere, & Minnema, 2005, 2006) as focused on Essential Competencies for Program Evaluators (ECPE). Their fit to an Asian country (Taiwan) was investigated by a two round Delphi technique with a small panel (*n* = 12) of experts. This led to minor adaptations to the framework.

The current effort reexamined that preliminary work with hundreds of respondents in Taiwan. Two scales (Likert, Fuzzy) were used to measure competencies via a needs format (desired and current statuses) which permitted the examination of discrepancies. By this means further evidence was obtained for the validity and generalizability of the competencies to the context in consideration. What was learned can also be used to enhance the professionalization of the field of evaluation. The major questions were:

(1) Do the modified competencies fit for a large sample of Taiwanese faculty members who are regularly involved in the evaluation (accreditation) of higher education?
(2) If it does, what are their perceptions of the "what should be" status and what is their own personal status? (Are needs apparent?)
(3) Are the results from instrument formats alike or different? (Was there convergent validity?)

2. Summary from the authors' prior 2011 research

The taxonomy of evaluator competencies of Stevahn et al. (2005) was reviewed by experts in Taiwan in a Delphi survey leading to small wording and content changes. Then, revised competencies in 6 clusters were put into a traditional needs assessment format and rated by the same panelists on Fuzzy importance and attainment scales. Ratings were high and the experts felt that the items and categories made sense for Taiwan. While this finding supported the relevance of Stevahn's framework to Taiwan, some unique skills were identified – 'Respects and follows the evaluation process' and 'Ensures the confidentiality of information' thus making for a more culturally sensitive investigation. For example, top-down or authoritarian environment, a strong link between outcomes and government policy, and the delicate interpersonal balance among the evaluators and those being evaluated were factors noted in panelist comments. By building in such factors, the assessment of evaluator skills, knowledge and dispositions would be improved.

Areas for change based on discrepancies from importance and current levels of competency, were not found. What might have contributed to the result (small panel size, issues in the Fuzzy

* Corresponding author. Tel.: +1 614 389 4585; fax: +1 614 688 3258.
 E-mail addresses: ivana@ntnu.edu.tw (Y.-F. Lee), Maltschuld1@columbus.rr.com (J.W. Altschuld), t83006@ntnu.edu.tw (L.-S. Lee).
 [1] Tel.: +886 2 7734 3398; fax: +886 2 2392 9449.
 [2] Tel.: +886 2 7734 3417; fax: +886 2 2392 1015.

scales, others) was discussed and further research with a larger sample and alternative scaling was recommended.

3. Evaluator roles and competencies

Evaluators deal with conflicts and complex conditions in practice as noted in the literature (Lynch, 2007; Mark, 2002; Montgomery, 2000; Patton, 2008; Russ-Eft, Bober, de la Teja, Foxon, & Koszalka, 2008; Turner, 2006). An especially important investigation was that done by Skolits, Morrow, and Burr (2009) in which they posited three phases of evaluation – pre, active, and post where evaluators served as managers, detectives, designers, negotiators, diplomats, researchers, judges, reporters, use advocates, and learners. They generated descriptions of each role and the kinds of competencies required.

Another effort was that of King, Stevahn and their colleagues. They perceived that tasks, skills, and content areas had to be "derived from a systemic process or validated by empirical consensus building among diverse professionals" (King et al., 2001, p. 230). To that end, they produced 61 competencies in six domains: systematic inquiry, professional practice, situational analysis, project management, reflective practice, and interpersonal competence (King et al., 2001; Stevahn et al., 2005). The categories and items in them should guide evaluators in reflection, self-analysis, and discussion about an array of knowledge, skills, and dispositions affecting practice.

Zorzi, McGuire, and Perrin (2002) used the framework for a study completed under the aegis of the Canadian Evaluation Society (CES). Forty-nine competencies in five practices were generated: reflective, technical, situational, management, and interpersonal practices. After member consultation and expert review, what they produced was approved by CES as related to its new credentialing system for evaluators (CES, 2009).

The efforts of all of the above researchers align well with Spencer and Spencer's (1993) model of competency design. Its steps are: (1) define performance effectiveness criteria; (2) identify a criterion sample; (3) collect data; (4) analyze data and develop a competency model; (5) validate the competency model, and (6) prepare applications of the competency model. Most of what was done previously tied into the first four steps but not so much 5 and 6, and especially the idea of validation (Wilcox, 2012). This presents an opportunity.

Wilcox (2012) explored this gap using a unified theory of validity from Messick (1989, 1995). Evidence was collected to demonstrate the extent to which the ECPE met five validity criteria: content-related, substantive-related, consequence-related, generalizability-related, and externally related evidence. For each criterion, the questions respectively were to what extent do the ECPE measure an evaluator's competence, are they inclusive of all necessary competencies for an evaluator to conduct evaluations, the use or interpretations of the competencies does not have negative consequences for evaluators, are they applicable to practice in various areas, and does competence correlate with other measures of competencies?

Data from a survey and interviews were gathered and analyzed indicating strong support for the first three criteria with mixed and limited evidence for the latter two. An additional concern is how well the competencies work across different contexts and samples (Stevahn et al., 2005; Wilcox, 2012). The current effort is exactly along these lines.

4. Setting and sample

The context was University Evaluation Programs administered by the Higher Education Evaluation and Accreditation Council (HEEAC) of Taiwan. The Council was established in 2005 to determine if degree-granting programs in a spectrum of fields in colleges and universities met standards of quality. The institutions must take part in the process once every five years except for engineering, computing, technical, and architectural education which are accredited by the Institute of Engineering Education. Accredited status has to be renewed periodically to demonstrate continuing fulfillment of HEEAC criteria.

Programs classified as conditionally accredited or failing may have their funding and/or the size of student enrollments reduced. With a high-stakes like these, the Council strove for consistency and fairness in its procedures (self-evaluation, site visits). Four to six professors or experts in appropriate disciplines are invited to be 'evaluators' for the visit. They have variable knowledge and experience in program management and evaluation and receive some training (usually one day) from HEEAC for the accreditation process (being on site, interviewing, compiling results, writing report). Most, even those who are considered specialists, have only part-time involvement in evaluation. 'Evaluators' in this setting differ from what would be the case in other parts of the world where the reference relates more to individuals professionally trained in the field.

Although the sample was drawn from those trained for site visits, the experience of some of its members was not limited to accreditation but extended to various other evaluation activities in higher education (see Section 6.1). Therefore, they would be capable of judging the competencies essential for a qualified program evaluator.

5. Methodology

Surveys with Likert and Fuzzy formats were generated about the aforementioned evaluator competency framework as adapted to Taiwan. Is it valid for the country in the eyes of hundreds of respondents and if so, are training needs uncovered? Does the structure of instruments produce similar or different results?

5.1. Instrument design

The instruments were based upon findings from the authors' 2011 study. They had six categories with 63 modified competencies from Stevahn's framework: Professional practice ($n = 11$), Systematic inquiry (18), Situational analysis (10), Project management (11), Reflective practice (5), and Interpersonal competence (8). Demographic and evaluation experience questions were included, e.g. how many times was a person involved in evaluation for higher education in the last 3 years, do you teach/conduct evaluation, do relevant research, etc.? An open-ended question was provided for comments.

Each item in a category was rated in terms of its importance for a qualified evaluator in Taiwan and the current competency of the respondent in regard to it. One form utilized a traditional 5-point Likert scale and the other a Fuzzy one. The Fuzzy scale has a range for rating instead of a single score going from above 0 to 1 in .1 increments, on the low and high ends respectively. Higher values denote more positive ratings (see Figs. 1 and 2).

5.2. Data analysis

Descriptive statistics (mean, standard deviation) were calculated for Likert scores. A critical value (3.5) was chosen arbitrarily by the researchers for a one-sample t-test, and if an item was significantly higher in importance it was designated an essential competency. Chen and Huang's (1992) method was employed for Fuzzy scores in which respondents' ratings were transformed to triangular Fuzzy numbers; then 'defuzzified' to synthesize the