Journal of MultiDisciplinary Evaluation
Volume 16, Issue 34, 2020

JMDE
Journal of MultiDisciplinary Evaluation

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

Refining and Measuring the Construct of Evaluative Thinking: An Exploratory Factor Analysis of the Evaluative Thinking Inventory

Jason S. McIntosh
Washington Elementary School District

Jane Buckley
JCB Consulting

Thomas Archibald
Virginia Tech

Background: Evaluative thinking has emerged as a key construct in evaluation, especially for evaluation practitioners and researchers interested in evaluation capacity building (ECB). Yet, despite increasing calls for more research on evaluation and, more specifically, for more research on ECB, until recently little empirical inquiry on the dimensions of evaluative thinking has been conducted.

Purpose: To address that lack, the purpose of the study presented in this paper is to refine the construct of evaluative thinking by exploring its underlying dimensions and to ascertain the internal consistency of an instrument developed to measure evaluative thinking, the Evaluative Thinking Inventory (ETI).

Setting: The ETI was developed as part of an ECB initiative focused on non-formal science, engineering, technology, and math (STEM) education in the United States, and was tested as part of a study focused on evaluating gifted education programs, also in the United States.

Intervention: Not applicable.

Research design: Survey research and exploratory factor analysis (EFA).

Data collection & analysis: The ETI was administered to participants in a study measuring the effectiveness of a tool used to conduct internal evaluations of gifted education programs. SPSS was used to conduct an EFA on 96 completed ETIs. Cronbach's alpha was used to estimate the internal consistency of the instrument.

Findings: The analysis of the ETI revealed a two-factor model of evaluative thinking (i.e., *believe in and practice evaluation* and *pose thoughtful questions and seek alternatives*). This study also provided internal consistency evidence for the ETI showing alpha reliabilities for the two factors ranging from 0.80 to 0.82. The ETI has potentially wide applicability in research and practice in ECB and in the field of evaluation more generally.

Keywords: *evaluative thinking; evaluation capacity building; research on evaluation; exploratory factor analysis.*

Introduction

Researchers and practitioners, especially those engaged in evaluation capacity building (ECB), agree that evaluative thinking is an important construct (Patton, 2018; Schwandt, 2018). King (2007), for example, described it as the ultimate outcome of ECB. However, until recently, there existed little agreement in the evaluation literature as to a formal definition of evaluative thinking, with even less guidance on how to measure it. It can be difficult to evaluate or research what you cannot define or accurately describe (Callahan, 1986). The recent volume of *New Directions for Evaluation* on the topic (Vo & Archibald, 2018) provided some welcomed empirical and conceptual clarity as to the contours of this somewhat elusive construct. For example, Vo, Schreiber, and Martin (2018) offer a conceptual model that consists of four key thematic domains (values, valuing, cognition, and application) based on their systematic literature review and conceptual analysis, while Fierro et al. (2018) conducted a multisite case study using document analysis, interviews, and a focus group to identify twenty-two indicators of evaluative thinking grouped into five categories—reflecting, perspectives, projecting, valuing evaluation, and use. But that 2018 volume did not propose a tool to measure evaluative thinking. The volume builds on Buckley, Archibald, Hargraves, and Trochim's (2015) previously proposed definition:

> Evaluative thinking is critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and making informed decisions in preparation for action. (p. 378).

Operationally defining evaluative thinking in this way allowed Buckley and Archibald (2011) to take the next logical step and create an instrument designed to measure it, entitled the Evaluative Thinking Inventory (ETI). That measure was developed and proposed based on Buckley and Archibald's conceptual research on the construct, as described in greater detail at the end of the following section. However, they stopped short of empirically testing the scale's quality, a task that we took up and moved forward in the study presented here.

As such, the purpose of the study presented in this paper was to answer two research questions: (1) What are the underlying dimensions of evaluative thinking? and (2) Does the ETI possess adequate internal consistency? The investigation of the ETI took place as a part of a larger study piloting a new program evaluation tool called the Depth and Complexity Program Evaluation Tool (McIntosh, 2015) designed to enable stakeholders to evaluate gifted education programs. This paper advances the theory and practice of ECB—responding to recent calls for more empirical research on ECB (Preskill, 2014; Suarez-Balcazar & Taylor-Ritzler, 2014; Wandersman, 2014)—by refining the components of evaluative thinking and providing initial reliability information on a measure of evaluative thinking. In addition, since evaluative thinking is increasingly seen as a concept that is fundamental to the whole field of evaluation (Vo & Archibald, 2018), the tool described in this paper has a potentially wide applicability in evaluation research and practice more generally.

Literature Review

ECB has steadily garnered increased attention in the field of evaluation over the past 20 years (Bourgeois & Cousins, 2013; Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012; Preskill & Boyle 2008; Stockdill, Baizerman, & Compton, 2002; Taylor-Powell & Boyd, 2008). ECB is frequently defined as "the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine" (Stockdill, Baizerman, & Compton, 2002, p. 14). A more recent definition of ECB frames it as "an intentional process to increase individual motivation, knowledge, and skills, and to enhance a group or organization's ability to conduct or use evaluation" (Labin et al., 2012, p. 308).

Recent years have witnessed a proliferation of conceptual models of ECB (Preskill & Boyle, 2008), including one that

was used as a guiding framework for a series of eight case studies of organizational evaluation capacity (Cousins, Goh, Elliott, & Bourgeois, 2014) and another developed through an extensive systematic review and synthesis of the ECB literature (Labin, 2014). Despite these advances, the need remains to advance the science, practice, and art of ECB by conducing more and better empirical research on ECB (Preskill, 2014). In this vein, there is a clear need to develop and apply better ways of measuring constructs associated with ECB, such as evaluative thinking. Furthermore, beyond the domain of ECB, evaluative thinking is increasingly seen as a key construct within the field of evaluation writ large (Patton, 2018), and as such is a salient phenomenon of interest to all who are interested in conducting research on evaluation, itself a growing area of focus within the evaluation field (Coryn et al., 2017).

The recipe for a quality evaluation requires a combination of evaluation know-how and evaluative thinking (Davidson, Howe, & Scriven, 2004). Without one the other suffers, resulting in decreased motivation, a tendency to resist change, and blind spots in perception (Buckley, Archibald, Hargraves, & Trochim, 2015). Humans are naturally inclined towards belief preservation, but this can be overcome by practicing critical thinking (Kahneman & Tversky, 1979; Lord, Ross, & Lepper, 1979). For that reason, evaluative thinking should be routinely and explicitly promoted and practiced (Buckley, Archibald, Hargraves, & Trochim, 2015). Being able to measure this construct then becomes very important—for instance, when an ECB facilitator wishes to ascertain the extent to which her ECB efforts are increasing the level of evaluative thinking over time with a group of individuals, or when an evaluation researcher wishes to better understand the relationship between evaluative thinking and other salient variables within a given organizational or programmatic context.

While there is a growing number of tools being developed and validated to measure evaluation capacity more broadly (Bourgeois, Toews, Whynot, & Lamarche, 2013; Gagnon, Aubry, Cousins, Goh, & Elliott, 2018; Preskill & Torres, 2000; Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry, Balcazar, 2013), to date, only two instruments have been developed to measure evaluative thinking. The first is called the Evaluative Thinking Assessment Tool and was created in 2005 by the Bruner Foundation. It was created as a result of a yearlong project known as the Evaluative Thinking in Organizations Study (ETHOS) designed to determine how evaluative thinking was related to effectiveness of organizations (Baker, Bruner, Sabo, & Cook, 2006).

The tool measures the degree to which organizations exhibit evaluative thinking in 15 core areas (e.g., mission, strategic planning, finance, governance, leadership, communications, client development). A percentage score for each area is generated by dividing the number of questions marked yes by the number of possible questions for each section. Although an interesting tool, no reliability or validity information has been published on it. It is also not appropriate for educational settings, the target population of the larger study conducted by McIntosh to pilot the Depth and Complexity Program Evaluation Tool (DC-PET).

As mentioned previously, the second tool is known as the Evaluative Thinking Inventory (Buckley & Archibald, 2011). The Evaluative Thinking Inventory (ETI) contains 20 questions measuring the degree to which participants: (a) pose thoughtful questions; (b) describe and illustrate thinking; (c) actively pursue deeper understanding; (d) express belief in the value of evaluation; and (e) seek alternatives (Buckley & Archibald, 2011). An example item from each construct is shown in Table 1. Individuals using the tool are presented with a six-point response scale and are asked to indicate the frequency with which they engage in the five constructs listed above. The tool is practical, easy-to-use, and very appropriate for educational settings.

Table 1
Example Items from Evaluative Thinking Inventory

Construct	Example Item
Posing thoughtful questions	I pose questions about assumptions and claims made by others.
Describing & illustrating thinking	I use models and/or other diagrams to clarify my thoughts.
Active engagement in the pursuit of understanding	I discuss evaluation strategies with my colleagues.
Seeking alternatives	I consider alternative explanations for claims.
Believing in the value of evaluation	I am eager to engage in evaluation.

The ETI tool was originally developed using accepted procedures for instrument design, described here. The ETI was developed in the context of a five-year intervention research project funded by the National Science Foundation and implemented by the Cornell Office for Research on Evaluation. The project was designed to develop, implement, and conduct research on a relational systems evaluation approach to ECB for non-formal STEM and other community-based educators in numerous contexts across the United States. Buckley and Archibald (2011) were involved in facilitating ECB workshops and offering evaluation coaching, as well as in collecting data on changes in evaluation capacity among participants. That practical experience prompted them to spontaneously begin to focus on the notion of evaluative thinking, which they then followed up on with a search of the evaluation literature, finding that, though sometimes mentioned and described, evaluative thinking had never previously been clearly defined. Based on their review of the evaluation literature, augmented by theoretical frameworks drawn from the fields of educational research, critical thinking, and cognitive science, they elucidated a concise initial theory- and experience-based definition of the construct, with five pillars or principles (Buckley, Archibald, Hargraves, & Trochim, 2015). The same accumulated base of experience and theory subsequently guided their brainstorming of potential survey items for each conceptual pillar. Then, through informal verification with evaluation experts and initial pilot testing with a few dozen ECB participants, the items were refined and face validity was established, yielding the version of the instrument used and tested in the present study.

Methods

In this study, the ETI was administered as a pre and post assessment to treatment and comparison group participants in a larger study measuring the effectiveness of a new tool used to conduct internal evaluations of gifted education programs (McIntosh, 2015). The Depth and Complexity Program Evaluation Tool (DC-PET) was developed to incorporate research and best practices from the fields of program evaluation, gifted education, and organizational change and includes both a workbook and web application. The Pre-K-Grade 12 Gifted Programming Standards (NAGC, 2010), the Program Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (Yarbrough, Shulha, Hopson, & Caruthers, 2011), and the 10 empowerment evaluation principles developed by Fetterman and Wandersman (2007) were also used as guidelines during the DC-PET's development. All participants were stakeholders of a gifted program in one of five states. A description of the sample is found in Table 2.

SPSS was used to conduct an exploratory factor analysis (EFA) on the 96 ETIs completed at the beginning of the study by both the treatment and comparison groups. Descriptive statistics were generated and the data were

cleaned accordingly. Principal axis extraction with direct oblimin factor rotation was then applied. Items loading <0.4 were suppressed. The data were then purposefully explored to look for patterns and noticeable trends.

Table 2
Description of Sample used in EFA[a]

	District											
	A	B	C	D	E	F	G	H	I	J	K	L
State	AZ	AZ	AZ	AZ	MN	AZ	AZ	AZ	AZ	GA	OH	SD
District Population	32,600	36,400	27,000	6,180	185	69	250	9,920	920	82,700	3,570	483
Designation	Suburb/ Public	Suburb/ Public	Suburb/ Public	Suburb/ Public	Rural/ BIE	Suburb/ Special School	Suburb/ Charter	Suburb/ Public	Rural/ Public	Suburb/ Public	Rural/ Public	Rural/ BIE
Free and Reduced	29%	17%	22%	64%	100%	N/A	N/A	34%	72%	38%	9%	100%
Participant Role *ns*												
Teachers	11	8	4	3	2	1	2	4	3	10	10	3
Parents	2	3	4	1	0	0	1	0	0	0	0	0
Admin.	1	2	2	3	2	1	1	0	1	0	1	0
Students	0	2	0	0	0	0	0	0	1	0	0	0
Other[b]	1	0	1	2	1	0	0	0	0	0	0	0
Total *n*	15	15	11	9	5	3	4	4	5	10	11	3

Note. [a]Statistics obtained from http://projects.propublica.org/schools/. [b]Includes school counselor, community member, and teacher coach.

Results

The first step in conducting the EFA was to clean the data. This involved filling in two pieces of data missing completely at random (MCAR) with the mean for the group and removing one multivariate outlier. The final descriptive statistics for each question on the ETI can be found in Table 3.

Table 3
Descriptive Statistics for Pre-Study Sample/Treatment & Comparison (*n* = 96)

Item	*M*	*SD*	Variance	Skewness	Kurtosis
Item 1	2.03	0.84	0.70	0.92	1.25
Item 2	2.17	0.66	0.44	-0.19	-0.70

Item	*M*	*SD*	Variance	Skewness	Kurtosis
Item 3	1.96	0.70	0.48	0.25	-0.26
Item 4	2.40	0.97	0.94	0.55	0.80
Item 5	1.55	0.66	0.44	1.02	0.88
Item 6	2.47	0.89	0.80	0.46	0.62
Item 7	2.21	0.71	0.50	0.76	1.96
Item 8	2.00	0.63	0.40	0.00	-0.43
Item 9	1.79	0.78	0.61	0.52	-0.74
Item 10	1.72	0.72	0.52	0.48	-0.95
Item 11	1.50	0.62	0.38	0.83	-0.29
Item 12	2.35	0.83	0.69	0.25	0.15
Item 13	1.99	0.79	0.62	0.41	-0.31
Item 14	2.08	0.76	0.58	0.15	-0.60
Item 15	1.80	0.80	0.65	0.99	1.56
Item 16	2.58	0.96	0.92	0.71	1.03
Item 17	1.96	0.66	0.44	0.05	-0.68
Item 18	2.41	1.05	1.11	0.72	0.59
Item 19	1.68	0.67	0.45	0.49	-0.74
Item 20	2.35	0.85	0.72	0.31	0.11

Research Question One

To reiterate, research question one was "What are the underlying dimensions of evaluative thinking?" Originally, the inventory was designed with five factors as described above. Four questions were created for each factor based on theory. When a five-factor model was forced using the principal axis extraction method and a direct oblimin factor rotation, the items did not correlate as expected. The initial factor loadings can be found in Table 4. In fact, there were numerous cross-loadings, questions that did not load on the hypothesized factor, and questions that did not load on any factor. Direct oblimin rotation was used due to the correlated nature of the factors (McCoach, Gable, & Madura, 2013).

Table 4
Original EFA Factor Loadings for Pre-Study Sample / Treatment &Comparison (*n* = 96)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
20. I enjoy discussing evaluation strategies with colleagues.					0.69
6. I discuss evaluation strategies with my colleagues.	0.61				0.32
16. I try to convince others that evaluation is important.	0.57				0.37
12. I articulate the relationship between my evaluation work and my intended claims.	0.74				

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
7. I articulate the logical justification of my evaluation strategy.	0.46				
10. I seek evidence for claims and hypotheses.		0.97			
9. I am wary of claims made by others without evidence to back them up.		0.74			
11. I am interested in understanding the logic behind things.		0.58			
14. I pose questions about assumptions and claims made by others.	0.45		0.36		
13. I reflect on assumptions and claims I make myself.	0.53			0.43	
17. I offer evidence for claims that I make.			0.31		
5. I take time to reflect about the way I do my work.					
4. I use models and/or other diagrams to clarify my thoughts.				0.99	
18. I use models and/or other diagrams to communicate my thinking to others.				0.79	
2. I am eager to engage in evaluation.	0.48				0.64
I describe my thinking to others.				0.56	
19. I believe evaluation is a valuable endeavor.					0.53
8. I consider alternative explanations for claims.			0.34	0.32	
15. I willingly make changes to the way I do my work.					

After purposefully exploring the data to look for patterns and using trial and error to delete questions one at a time, a noticeable trend emerged. Certain items within the *posing thoughtful questions* factor and the *seeking alternatives* factor were loading on the same factor instead of separately. Also, items within the *expressing belief in the value of evaluation* factor and the *actively pursuing deeper understanding* factor were loading on the same factor instead of separately. The ties between these constructs made logical sense and so a three-factor model was run. The new factors were renamed: (a) *describing and illustrating thinking*; (b) *posing thoughtful questions and seeking alternatives*; and (c) *believing in and practicing evaluation*.

The new 14 item three factor model, with factor loadings <0.4 suppressed, is shown in Table 5. The path diagram is shown in Figure 1 and alpha reliabilities with descriptive statistics by factor are shown in Table 6. Together, the three factors accounted for 49.68% of the variance. The correlation matrix for the final model is shown in Table 7.

Table 5
Final EFA Factor Loadings for Pre-Study Sample/Treatment and Comparison (n = 96)

Item[a]	Factor 1	Factor 2	Factor 3
20. I enjoy discussing evaluation strategies with colleagues.	0.88		
2. I am eager to engage in evaluation.	0.68		
16. I try to convince others that evaluation is important.	0.61		
19. I believe evaluation is a valuable endeavor.	0.57		
6. I discuss evaluation strategies with my colleagues.	0.56		
14. I pose questions about assumptions and claims made by others.		0.68	

Item[a]	Factor 1	Factor 2	Factor 3
17. I offer evidence for claims that I make.		0.65	
8. I consider alternative explanations for claims		0.61	
9. I am wary of claims made by others without evidence to back them up.		0.55	
13. I reflect on assumptions and claims I make myself.		0.54	
3. I suggest alternative explanations and hypotheses.		0.52	
5. I take time to reflect about the way I do my work.		0.41	
4. I use models and/or other diagrams to clarify my thoughts.			0.92
18. I use models and/or other diagrams to communicate my thinking to others.			0.84

Note. [a]Items loading <0.4 were suppressed.

Table 6
Evaluative Thinking Inventory Descriptive Statistics by Factor for Pre-Study Sample (*n* = 96)

		Response Percentage										
Factor	Item	1	2	3	4	5	6	*M*	*SD*	r^a	α if item removed	α
Believing in and practicing evaluation	2	15	54	31	0	0	0	2.17	0.66	0.55	0.80	0.82
	6	12	40	40	5	3	0	2.47	0.89	0.54	0.80	
	16	9	42	34	12	2	1	2.58	0.96	0.66	0.77	
	19	44	45	11	0	0	0	1.68	0.67	0.60	0.78	
	20	15	44	34	6	1	0	2.35	0.85	0.72	0.74	
Posing thoughtful questions and seeking alternatives	3	25	55	19	1	0	0	1.96	0.69	0.55	0.77	0.80
	5	53	40	6	1	0	0	1.55	0.66	0.50	0.78	
	8	20	60	20	0	0	0	2.00	0.63	0.55	0.77	
	9	42	38	19	1	0	0	1.79	0.78	0.40	0.80	
	13	28	48	21	3	0	0	1.99	0.79	0.56	0.77	
	14	23	48	27	2	0	0	2.08	0.76	0.60	0.76	
	17	24	56	20	0	0	0	1.96	0.66	0.59	0.76	
Describing and illustrating thinking	4	18	39	32	10	0	1	2.40	0.97	0.80	0.85	0.89
	18	15	44	34	6	1	0	2.41	1.05	0.80	0.84	

Note. [a]With corrected item total.

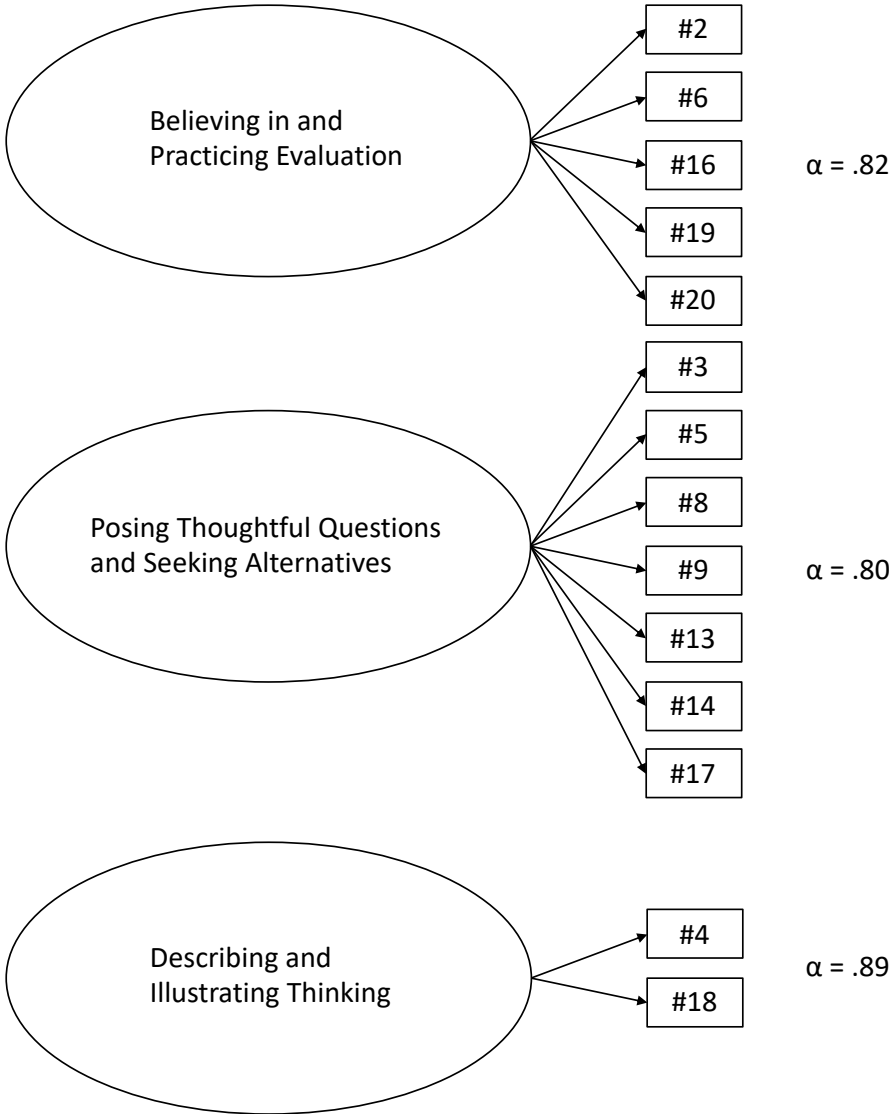

Figure 1. *Path diagram for the Evaluative Thinking Inventory after EFA.*

Table 7
Correlation Matrix for Pre-Study Sample/Treatment & Comparison (*n* = 96)

	Item 2	Item 3	Item 4	Item 5	Item 6	Item 8	Item 9	Item 13	Item 14	Item 16	Item 17	Item 18	Item 19	Item 20
Item 2	1													
Item 3	0.31	1												
Item 4	-0.07	0.28	1											
Item 5	0.17	0.35	0.31	1										
Item 6	0.30	0.24	0.31	0.31	1									
Item 8	0.28	0.50	0.17	0.28	0.21	1								
Item 9	0.15	0.30	0.01	0.25	0.04	0.3	1							
Item 13	0.23	0.33	0.21	0.41	0.40	0.4	0.24	1						
Item 14	0.10	0.36	0.14	0.35	0.22	0.3	0.38	0.51	1					
Item 16	0.39	0.31	0.01	0.25	0.51	0.2	0.25	0.44	0.41	1				
Item 17	0.23	0.43	0.14	0.32	0.30	0.5	0.23	0.46	0.49	0.39	1			
Item 18	0.01	0.28	0.80	0.32	0.38	0.2	0.05	0.27	0.35	0.21	0.27	1		
Item 19	0.53	0.31	0.09	0.31	0.32	0.3	0.27	0.35	0.28	0.54	0.37	0.20	1	
Item 20	0.57	0.35	0.08	0.23	0.53	0.2	0.13	0.26	0.25	0.56	0.27	0.22	0.51	1

One issue still remained. Factor three now only had two questions that are 0.4 or greater. Standard practice dictates that a factor must have a minimum of three questions (Brown, 2006). The results of the EFA were presented to the creators of the inventory, Buckley and Archibald (2011), through a teleconference and an agreement was made to create and pilot additional questions for the *describing and illustrating thinking* factor at a later date.

Research Question Two

As a reminder, research question two was, "Does the ETI possess adequate internal consistency?" The alpha-reliability estimates of the data for internal consistency were 0.82 for factor one, 0.80 for factor two, and 0.89 for factor three. A Cronbach's α greater than 0.70 is considered adequate for an affective measure (McCoach, Gable, & Madura, 2013).

Discussion

The analysis of the Evaluative Thinking Inventory revealed a two-factor model of evaluative thinking (i.e., *believe in and practice evaluation* and *pose thoughtful questions and seek alternatives*). Evidence for the third factor, *describing and illustrating thinking*, was not sufficient to draw conclusions, leaving room for additional work to further develop, refine, and test the construct in the future. This study also provided internal consistency evidence for the ETI showing alpha reliabilities for the two factors ranging from 0.80 to 0.82. The resulting 18 item ETI, which includes pilot questions for a possible third factor, is included in the Appendix A to promote its wider use and further testing in the field.

Future Research

The next logical step is to collect a larger number of completed inventories and pilot new questions for factor three. A list of possible new questions for factor three includes the following:

- I use diagrams and/or illustrations to clarify my thoughts

- I use diagrams and/or illustrations to communicate my thinking to others
- Diagrams and/or illustrations help me think about ideas
- I describe the thinking behind my work to my colleagues
- I describe the thinking behind my decisions to others
- I brainstorm with colleagues to develop plans and/or ideas

The new version of the ETI could be administered to participants similar to those in the original study, but also to those outside the field of education. The results could then be compared to look for similarities and differences.

A second future study to be undertaken is the collection and analysis of additional completed instruments to perform confirmatory factor analysis (CFA). Confirmatory factor analysis is used to determine if a hypothesized model is consistent with the data. A preliminary CFA was conducted on the post data used in this study, but since it is not accepted practice to use the same sample for both EFA and CFA, that initial CFA is not shared here. No single study stands alone and any attempt to hasten the progress of a discipline by attempting to use a single data set to identify a construct and address validity is not advisable (Hurley, Scandura, Schriesheim, Brannick, Seers, Vandenberg, & Williams, 1997). A third possible study relates to the closeness in the relationship between evaluative thinking and critical thinking. It would be interesting to know if evaluative thinking, as measured by a pre and post administration of the ETI, would increase following several weeks of daily required critical thinking exercises.

Limitations

The first limitation to this study is the sample size. A large number of participants is always desirable when analyzing quantitative data in particular, which is especially true when seeking to better understand the structure of a latent variable or construct. However, the sample size in this study is above the lower threshold of acceptable sample sizes for EFA (de Winter, Doudou, & Wieringa, 2009). A second limitation to the study is that both traditional EFA and final model reliability estimates using Cronbach's alpha tend to introduce some error when applied to ordinal rather than continuous data (Sanders, Gugiu, & Enciso, 2015), as in the case of the present study. A third limitation is that all participants were parents or educators associated with a gifted program, which may limit the generalizability of the findings to the broader population. A fourth potential limitation is that the ETI is based on a definition of ET that does not explicitly prioritize valuing and reasoning, per se, which some scholars have more recently suggested are key elements of evaluative thinking (Schwandt, 2018; Vo, Schreiber, & Martin, 2018). A final notable limitation is that a majority of individuals chose responses 1-3 when responding to each statement (Table 6), thereby limiting the variability of the data. A lower rating corresponds to a higher self-reported frequency of participating in that action. We can only hypothesize as to why this is the case (e.g., educators routinely evaluate students and may therefore conflate that with evaluating programs, respondents not knowing what they do not know). Regardless of the reason, the skewedness of responses warrants further examination, and may have impeded our ability to accurately elucidate the underlying dimensions of the construct. Despite these limitations, we posit that this study does indeed contribute to the field's understanding of the construct of ET, and yields an instrument with adequate internal consistency to warrant further use and testing in a wider array of evaluation and ECB contexts.

Conclusions

Evaluative thinking has emerged as a key construct in evaluation, especially for those evaluation practitioners and researchers interested in ECB. Yet, despite increasing calls for more research on evaluation and, more specifically, for more research on ECB, until recently few empirical assessments on the dimensions of evaluative thinking had been conducted. To address that lack, this paper presented the results of a study that refined the construct of evaluative thinking—

identifying a two-factor model of evaluative thinking (i.e., *believe in and practice evaluation* and *pose thoughtful questions and seek alternatives*)—and provided initial consistency evidence for a tool designed to measure evaluative thinking known as the ETI. The results of the study presented above suggest that the modified 18 item ETI is a high-quality measure of evaluative thinking that can be used for research and evaluation on ECB and as a tool for research on evaluation in diverse contexts.

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 0814364.

References

Baker, A., Bruner, B., Sabo, K. & Cook, A. (2006). *Evaluation capacity and evaluative thinking in organizations*. Retrieved from: http://www.evaluativethinking.org/docs/EvalCap_EvalThink.pdf

Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation. 34*(3), 299-319.

Bourgeois, I., Toews, E., Whynot, J., & Lamarche, M. K. (2013). Measuring organizational evaluation capacity in the Canadian federal government. *Canadian Journal of Program Evaluation*, *28*(2), 1-19.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Buckley, J. & Archibald, T. (2011). The Evaluative Thinking Inventory. *Cornell University*. Retrieved from: https://cornell.qualtrics.com/jfe/form/SV_eLKEC3flFDNuU8A

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

Callahan, C. (1986). Asking the right questions: The central issue in evaluating programs for the gifted and talented. *Gifted Child Quarterly, 30*, 38-42.

Coryn, C. L., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., ... & Schröter, D. C. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation*, *38*(3), 329-347.

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

Davidson, E., Howe, M., & Scriven, M. (2004). Evaluative thinking for grantees. In M. Braverman, N. Constantine, & J.K. Slater (Eds.), *Foundations and evaluation: Contexts and practices for effective philanthropy* (pp. 259-280). San Francisco, CA: Jossey-Bass.

de Winter, J. D., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147-181.

Fetterman, D. & Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York: Guilford Press.

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

Gagnon, F., Aubry, T., Cousins, J. B., Goh, S. C., & Elliott, C. (2018). Validation of the evaluation capacity in organizations questionnaire. *Evaluation and Program Planning*, *68*, 166-175.

Hurley, A., Scandura, T., Schriescheim, C., Brannick, M., Seers, A., Vandenberg, R., & Williams, L. (1997). Exploratory and confirmatory factor analysis guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18*, 667-682.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation, 116*, 45-59.

Labin, S. N. (2014). Developing common measures in evaluation capacity building: An iterative science and practice process. *American Journal of Evaluation*, *35*(1), 107-115.

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

Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology,* 37, 2098-2109.

McCoach, B., Gable, R., & Madura, J. (2013*). Instrument development in the affective domain: School and corporate applications.* New York: Springer.

McIntosh, J. (2015). *The depth and complexity program evaluation tool: A new method for conducting internal evaluations of gifted education programs* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

NAGC. (2010). Pre-K-Grade 12 gifted programming standards. Retrieved from: http://www.nagc.org/uploadedFiles/Information_and_Resources/Gifted_Program_Standards/K-12%20programming%20standards.pdf

Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, *35*(1), 116-119.

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

Preskill, H., & Torres, R. T. (2000). Readiness for organizational learning and evaluation instrument. Available from H. Preskill, hallie.preskill@fsg.org.

Sanders, M., Gugiu, P. C., & Enciso, P. (2015). How good are our measures? Investigating the appropriate use of factor analysis for survey instruments. *Journal of Multidisciplinary Evaluation*, *11*(25), 22-33.

Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *Evaluative Thinking. New Directions for Evaluation.158*, 125-137.

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. In D. W. Compton, M. Baizerman, & S. H. Stockdill (Eds.), *The Art, Craft and Science of Evaluation Capacity Building: New Directions for Evaluation, 93*, 7-25.

Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2014). Moving from science to practice in evaluation capacity building. *American Journal of Evaluation*, *35*(1), 95-99.

Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation, 120,* 55-69.

Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D., Balcazar, F. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation, 34*(2), 190-206.

Vo, A. T., & Archibald, T. (2018). New directions for evaluative thinking. In A. Vo & T. Archibald (Eds.), *Evaluative Thinking: New Directions for Evaluation, 158*, 139-147. doi.org/10.1002/ev.20317

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

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

Yarbrough, D., Shulha, L., Hopson, R., & Caruthers, F. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. Thousand Oaks, CA: Sage.

Appendix A

Items on the revised Evaluative Thinking Inventory (ETI).

Response options: Very Frequently; Frequently; Occasionally; Rarely; Very Rarely; Never.

1. I discuss evaluation strategies with my colleagues.
2. I am eager to engage in evaluation.
3. Diagrams and/or illustrations help me think about ideas.
4. I am wary of claims made by others without evidence to back them up.
5. I describe the thinking behind my decisions to others.
6. I take time to reflect on the way I do my work.
7. I try to convince others that evaluation is important.
8. I consider alternative explanations for claims.
9. I brainstorm with colleagues to develop plans and/or ideas.
10. I believe evaluation is a valuable endeavor.
11. I use diagrams and/or illustrations to clarify my thoughts.
12. I suggest alternative explanations and hypotheses.
13. I reflect on assumptions and claims I make myself.
14. I pose questions about assumptions and claims made by others.
15. I enjoy discussing evaluation strategies with colleagues.
16. I describe the thinking behind my work to my colleagues.
17. I offer evidence for claims that I make.
18. I use diagrams and/or illustrations to communicate my thinking to others.