

Developing and Validating the Claremont Evaluative Thinking Scale

Rebecca Polivy
Claremont Graduate University

Stewart I. Donaldson
Claremont Graduate University

Journal of MultiDisciplinary Evaluation
Volume 22, Issue 52, 2026

JMDE
Journal of MultiDisciplinary Evaluation

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

Background: Evaluative thinking (ET) has increasingly been recognized as an essential ability for successful evaluation practice, evaluation capacity building, engaged employees, and even healthy democracies. However, the relationship between increased capacity to think evaluatively and its potential outcomes has not been sufficiently empirically demonstrated. To better understand the impact and value of thinking evaluatively, it behooves us to find innovative ways to measure the construct of evaluative thinking.

Purpose: This study enhances our understanding of, and capacity to measure, individual-level dispositions of evaluative thinking, laying the groundwork for future research into understanding its impact more deeply.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: Mixed Methods.

Data Collection and Analysis: Interviews, Focus Groups, Exploratory Factor Analysis.

Findings: Factor analysis revealed six factors with strong item loadings, and a Cronbach's alpha test confirmed reliability ($\alpha = 0.96$). The six dimensions of ET probed with this scale are *considers context, seeks multiple perspectives, intentional/planful, poses thoughtful questions, constructs sound arguments, and engages in reflective practice.*

Keywords: *Evaluative thinking; research on evaluation (RoE), evaluation theory; evaluation practice*

Evaluative thinking (ET) has increasingly been recognized as an essential ability for successful evaluation practice, evaluation capacity building, engaged employees, and even healthy democracies. However, the relationship between increased capacity to think evaluatively and its potential outcomes has not been sufficiently empirically demonstrated. Without empirical research exploring these relationships, evaluation scholars, practitioners, and capacity builders, are left to wonder what, if any, is the real value of building the capacity of individuals to think evaluatively. What impact, if any, does evaluative thinking *really* have on evaluation processes, organizational performance, and society at large? To better understand the impact and value of thinking evaluatively, it would behoove us to find innovative ways to measure the construct of evaluative thinking.

A reliable and valid scale offers one opportunity to test what impact, if any, evaluative thinking has on evaluation processes and teams and, possibly, society. Additionally, such a scale, especially when coupled with other methodological tools, could unearth other predictive qualities of thinking evaluatively. If, indeed, evaluative thinking contributes to better evaluation and organizational outcomes, it is worth investing more time and financial resources to building evaluative thinking capacity. If greater capacity to think evaluatively has no significant impact on other aspects of individual or organizational life, then perhaps resources are better placed elsewhere. To answer these larger questions, we must first understand how to measure ET.

Background

While the idea that evaluation, or the process of coming to a judgment of value, requires a unique way of thinking and reasoning has underlain the earliest writings in the field, the term “evaluative thinking” (ET) has been becoming more prevalent in the evaluation literature in the past decade (Patton, 2018; Vo, 2013). Scholars in the field of evaluation advocate that building capacity to think evaluatively is beneficial, not just for professional evaluators, but amongst non-evaluators as well. Buckley and colleagues (2015) believe it will make for more successful evaluation processes and thus more impactful programs. Patton (2018) states that ET is not only essential for evaluation practice but is a life-or-death imperative for the times in which we are living (Patton, 2018; Vo & Archibald, 2018). Patton (2018) believes ET can be a weapon against the dangers of widespread misinformation

consumption and the threats of confirmation bias—both societal ills prevalent today.

What is Evaluative Thinking (ET)?

In essence, “ET” refers to the cognitive approach those engaging in evaluative activities bring to reasoning about programs, policies, and other evaluands (i.e., subjects of evaluations). Early conceptualizations cast ET largely as critical thinking applied in evaluation contexts. For example, Scriven (1991) argued that core ET skills include critical thinking, logical reasoning, and understanding others’ perspectives. Vo (2013), through a consensus-building approach in which perspectives from highly regarded evaluators were gathered, defined ET as

... a particular kind of critical thinking and problem-solving approach that is germane to the evaluation field. It is the process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgements in a transparent fashion. (p. 107)

Buckley and colleagues (2015) define ET as

... 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 informing decisions in preparation for action. (p. 378)

Buckley and colleagues urged the field to draw on the rich literature of analytical reasoning and evidence use developed in critical thinking research. This foundational work established a base for understanding ET as a skill set centered on inquisitive, evidence-based, and analytical thinking.

Broadening the Conceptual Foundation of ET

More recent scholarship, however, cautions against a narrow or purely technocratic view of ET, arguing that it is not merely an individual cognitive skill but is a dynamic, collaborative, and values-driven process embedded in social contexts. Schwandt (2018) and Gates (2018), for instance, argue that ET should be understood as a social practice, not just an isolated mental exercise. They portray ET as

an approach to inquiry and learning that is inherently relational and situational, shaped by interest holder dialogues, attentive to power dynamics, and responsive to cultural context. Furthermore, Schwandt (2018) emphasizes, ET is unique in that it deliberately integrates values with evidence in a systematic way to inform judgments of merit and worth. In other words, ET follows a “procedural logic” of evaluation that combines empirical data with explicit criteria and interest holder values to reach transparent evaluative conclusions. Such integration of facts with values situates ET as a deeply values-driven process, reinforcing that evaluative judgments are never value-neutral but rest on negotiated standards of what is good or important. This broader lens reveals that ET encompasses not only logical analysis, but also ethical reasoning, individual values, interpersonal considerations, and systemic and contextual awareness.

Relational, Ethical, and Systemic Dimensions

Embracing a broader view of ET means acknowledging its relational, systemic, and ethical dimensions alongside the technical. Relationally, ET involves understanding diverse perspectives and fostering collaborative inquiry. Scriven (1991) noted the importance of understanding others—how they think, reason, judge, interact (pp. 1-43) as a core element of ET. Similarly, Hazel Symonette’s (2004) work on culturally responsive evaluation calls for diversity-grounded, equity-minded evaluative thinking/doing/being that engages empathic perspective-taking across different social positions. In Symonette’s view, evaluative thinkers must cultivate self-awareness (“self-to-self”), empathy with others (“self-to-others”), and awareness of broader systems (“self-to-systems”) to conduct evaluations that are socially just and contextually valid. Ethically, ET demands continuous reflection on one’s values and biases and the implications of one’s actions and judgments in given contexts. Tovey and Archibald (2024) connect ET with reflective practice and practical wisdom, suggesting that effective evaluative thinkers regularly reflect on their practice to align their decisions with ethical principles and situational demands. This stance ensures that ET is not a rote application of tools but rather a mindful, morally and contextually attentive process of reasoning. Moreover, evaluative thinkers are attentive to systemic and contextual factors. They probe beyond immediate data to examine unintended consequences, complex interactions, and the broader environment in which a program

operates. Contemporary discussions highlight complexity and systems science as key informants of ET, stressing that evaluators must think systemically to address “wicked” social problems (Gates, 2018; Patton, 2018). Together, these relational, ethical, and systemic dimensions position evaluative thinking as a holistic mode of inquiry—one that is critical and analytical, yet also context-aware, collaborative, and grounded in values and ethics.

How ET Connects to Related Constructs

Given this rich, multifaceted view, ET can be seen as an intersecting construct that overlaps with—yet remains distinct from—several other well-known forms of thinking. For the most part, there is agreement that ET has its roots in critical thinking. However, there is discussion about whether ET is critical thinking applied to evaluation contexts (Buckley et al., 2015) or if it goes beyond just critical thinking to incorporate other types of thinking and dimensions (Lai, 2011). Patton (2018) suggests ET goes beyond critical thinking to also include creative, practical, and inferential thinking. Vo and colleagues (2018) conclude:

Making reasoned, evidence-based choices set critical thinking and evaluative thinking apart from judgements made based only on deeply held and unchallenged beliefs (e.g., stereotypes). Making a reasoned choice about value and being able to defend it is what distinguishes evaluative thinking from critical thinking. (p. 40)

The literature to date points to the distinction between critical thinking and ET as involving context and values-based judgments.

Systems thinking intersects with ET in its focus on context and interrelationships; evaluative thinkers, like systems thinkers, seek to understand the bigger picture and the dynamic interactions that affect program performance. In fact, ET is often needed to make sense of complexity, as it combines analytical rigor with a holistic outlook to determine what works for whom and under what conditions. Ethical reasoning is inherent to ET because evaluation inherently involves determining what is valuable and for whom. As such, ET requires grappling with questions of values, equity, and responsibility, much like ethical reasoning, to ensure that conclusions are justifiable and culturally respectful. Reflective practice is both a cousin to and a component of ET; both involve learning from experience and questioning one’s

assumptions. Indeed, engaging in evaluative thinking typically means adopting a reflective stance—continually asking “Why did this result occur? How do my own assumptions influence my interpretation? What can be improved?” (Tan, Kocsis, & Burry, 2023 c.f. Schön, 1983). This reflexivity links ET to concepts of metacognition, as evaluative thinkers must be aware of their own cognitive processes and biases, monitoring how they (and others) are reasoning through an evaluative problem. Finally, information literacy—the ability to find, assess, and use information effectively—is a prerequisite for strong ET. By definition, ET involves the “use of data and evidence in argumentation” and an “attitude of inquisitiveness” that drives one to seek out relevant information (Buckley et al., 2015). Without sound information literacy skills, one cannot gather credible evidence nor critically appraise it to support evaluative judgments. In summary, ET draws on these related constructs (critical, systemic, ethical, reflective, and metacognitive thinking, as well as information literacy) to form a unique synthesis suited to evaluative inquiry. It is this synthesis—applying critical reasoning and values reflection in a specific context to arrive at a sound judgment—that makes ET a distinct and essential concept in both formal and informal evaluation.

Evaluative Thinking as Added Value?

Thinking evaluatively is considered a lynchpin for building individual and organizational evaluation capacity. Evaluation capacity building (ECB) is defined as “the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine” (Stockdill et al., 2002, p. 105). Archibald, Buckley, Hargraves and their various teams (2015, 2016, 2021) theorized that building the capacity for evaluators and non-evaluators to think evaluatively improves evaluation processes and success. Having a group of people—rather than a single evaluator—who feel encouraged to ask thoughtful questions about what is working and what can be improved, who understand their role in an evaluation, and who share in reflective conversations about their work is considered a key ingredient for developing a culture of learning. However, there is no body of empirical evidence to back up this claim.

A case needs to be made for why the return on an investment in ECB would be greater than that of another professional development opportunity or investing in an external evaluator. Leviton (2014) poses the questions, “Once an organization begins to ‘think evaluatively,’ what follows in the longer term?” (p. 91). What would be the added investment and subsequent added value, for an organization, of building the capacity of multiple staff members to support an evaluation?

While not overtly stated, qualitative evidence from studies on the effectiveness of building ET capacity allude to the possibility that its impact could go beyond just supporting quality evaluations. Participants who were part of efforts to build ET capacity reported feeling better equipped to have difficult conversations with superiors. Members of teams retold stories of, together, moving from being stuck to unstuck (Chauveron et al., 2021, Hargraves et al., 2021). These examples imply that increased capacity to think evaluatively fosters better communication, teamwork, feelings of belonging, and ownership of their work. In the era of the Great Resignation,¹ when places of business are rethinking what it takes to retain employees, could building the capacity of individuals to think evaluatively offer organizational value above and beyond successful evaluations?

Can Evaluative Thinking Be Measured?

Despite consensus that evaluative thinking is crucial for effective evaluation and organizational learning, the field still lacks well-validated tools to measure ET reliably. A clear conceptual foundation, as outlined above, is not merely academic; it also guides what we measure and why. A body of evidence exploring the value of increasing people’s capacity to think evaluatively could be instrumental in making the case for ECB initiatives, which, in turn, could enhance both the practice and the field of evaluation and quite possibly society at large. To do so, there must be a way to measure ET to make note of where it exists and how it manifests and understand its predictive qualities.

Developing a robust evaluative thinking scale serves several important purposes. First, such a scale can help evaluate the effectiveness of training and capacity-building efforts in the evaluation field. Many evaluation capacity-building programs aim to foster ET among practitioners and program staff,

¹ “The Great Resignation” was the name given to the economic trend in America starting in 2021, post pandemic, when employees were quitting their jobs at

unusually high rates in search of better wages, conditions, and workplace culture (https://en.wikipedia.org/wiki/Great_Resignation).

yet without a valid measure it is difficult to gauge progress or identify areas needing improvement. Second, an ET scale could be used in professional development and education; for example, graduate programs in evaluation or public administration could assess students' growth in evaluative thinking skills, or organizations could use the scale to identify strengths and gaps in their teams' evaluative mindsets.

Indeed, as Patton (2018) and Buckley et al. (2015) suggest, the potential audience for an evaluative thinking measure goes beyond professional evaluators. Managers, educators, policy analysts, and non-profit leaders—anyone engaged in making evidence-informed decisions—could benefit from strengthening and assessing ET capacities. Prior research in fields like education suggests that ET underpins effective practice (e.g., teachers using ET to continually improve student learning outcomes). This implies that an ET scale might find applicability in diverse sectors as a tool for promoting reflective, evidence-based decision making. Third, articulating and measuring ET can advance theory and scholarship. A validated scale grounded in the multidimensional conceptualization of ET will allow researchers to study how ET relates to other constructs (such as those discussed above) and to outcomes like evaluation use or program success. It also enables empirical testing of assumptions in the literature—for instance, is ET truly enhanced through certain training approaches? Does stronger ET correlate with more adaptive and learning-oriented organizations? These questions can only be answered with sound measurement.

Currently two scales exist to measure ET. The [Evaluative Thinking Assessment Tool](#), created in 2005 by the Bruner Foundation, measures ET at the organizational level but has not been validated nor peer-reviewed. The Evaluative Thinking Index (ETI), created by Buckley and Archibald (2011), measures ET on the individual level. The ETI has been peer-reviewed and is deemed valid in probing

two out of five hypothesized factors of ET: (1) believing in and practicing evaluation and (2) posing thoughtful questions and seeking alternatives (McIntosh et al., 2020). However, other factors hypothesized by Buckley and Archibald, such as describing and illustrating thinking, were not sufficiently probed by the items on their scale (McIntosh et al., 2020). Additionally, “valuing” was not included in the ETI’s working definition of ET, even though “valuing” has been deemed an important element of ET (McIntosh et al., 2020). Furthermore, while it has been judged appropriate for different contexts, the ETI’s only published study has been its use amongst evaluators in an education setting.

Neither of the two scales has been used widely to build a body of evidence about the contributions ET can make to the field of evaluation and beyond. This study built off the work of the ETI to create a scale to measure ET that is reliable and valid, that probes its various individual-level dimensions, and that can be used in a variety of contexts, including both in- and outside of formal evaluation contexts. Our goal is to provide scholars and practitioners with a practical tool that not only starts to measure dimensions of evaluative thinking but also promotes deeper understanding of this construct by validating its dimensions while also acknowledging the limitations of measurement. By attempting to firmly anchor the scale in a comprehensive theoretical foundation, one that integrates technical skills and ethical praxis, individual and systemic thinking, critical analysis and reflective insight, we aim to progress beyond previous conceptual limitations and contribute to the evolution of evaluative thinking as both a field of study and a practice. Ultimately, this work seeks to support evaluators and change agents in cultivating genuine evaluative thinking, thereby enhancing the quality and impact of evaluation in service of learning and improvement and possibly even democracy.

Table 1. Existing Scales to Measure Evaluative Thinking

Measure	Purpose	Dimensions	Validation
Evaluative Thinking Assessment Tool	To assess the extent to which evaluative thinking is present in various organizational capacity areas.	• Asking questions of substance • Determining what data are needed to address the questions • Gathering appropriate data in systematic ways • Analyzing data and sharing results • Developing strategies to act on evaluation findings.	None
Evaluative Thinking Index	To measure evaluative thinking in individuals.	• Believing in and practicing evaluation • Posing thoughtful questions and seeking alternatives^a	A two-factor model for ET was validated and considered reliable for measuring ET.
CETS	To measure ET in individuals, including both evaluators and non-evaluators.	• Considers context • Seeks multiple perspectives • Intentional/planful • Poses thoughtful questions • Constructs sound arguments • Engages in reflective practice	Face, content, and construct validity and reliability

^a While the Evaluative Thinking Index originally probed five conceptualized factors, the validation process collapsed it to two valid factors (McIntosh et al., 2020)

Methodology

As the goal of this study is to develop a quantitative instrument to ultimately support the field in understanding ET and its relationships with other constructs, this study used an exploratory sequential mixed-methods approach with the goal of acknowledging and surfacing the rich theoretical understanding of ET through qualitative approaches and attempting to strengthen and validate these conceptualizations using quantitative methods (Creswell & Plano Clark, 2017). Additionally, this approach was employed for triangulation purposes and to support and strengthen any findings of the study (Maxwell, 2013). Together, this mixed-methods approach aimed to deepen and validate our understanding of ET. Stage 1 of the study, the construct clarification stage, consisted of qualitative methods of literature review and expert interviews. Stage 2 of the study, the item development and refinement stage, consisted of focus groups of emerging and non-evaluators and a survey administered to 250 people whose responses were subject to quantitative analysis for the purposes of data reduction and to test for reliability and validity.

Stage 1: The Dimensions of ET

Stage 1 aimed to answer the research question: What dimensions demonstrate or help us recognize ET?

Literature Review. Twenty-seven peer-reviewed articles were read to explore the construct of ET and past attempts to define it. These articles were located using keywords such as “evaluative thinking,” “research on evaluative thinking,” “thinking evaluatively,” and “measuring evaluative thinking” to search both evaluation journals as well as journals in the fields of education and leadership.

Interviews. Through the analysis of the literature, a refined list of key dimensions related to ET were drafted into a semi-structured interview protocol. This protocol included open-ended questions asking for expert opinions on aspects of the leading definitions of ET which remained unclear, such as the difference between ET and evaluative practice and whether the experts thought ET only occurs in an evaluation context.

Interviews were conducted with a non-random, purposive sample of nine “expert” scholars on the topic of evaluative thinking. To have expertise, meant having at least 10 years of research, teaching,

and practical experience in the field of evaluation and/or having written explicitly, or made implied comments, about ET in their publications. While the interviews were conducted with experts as defined above, the line of questioning aimed to elicit their thoughtful, informed opinions on what unique dimensions of ET might exist both in and outside of formal evaluation contexts. The interviews were conducted and recorded via Zoom and transcribed via the online platform Otter.ai. Qualitative coding was conducted using thematic content analysis.

Stage 2: Item Generation and Refinement

Stage 2 focused on item development and refinement and was guided by the following research questions: (1) Is there internal consistency among these items? (2) Do items probing these ET dimensions have validity?

Item Generation. Key factors of ET emerged from Stage 1's findings, and a list of 40 items that operationalize and triangulate those factors was developed. The items attempted to meet the criteria of ease and consistency of interpretability and required Likert-type responses on a scale of 1 to 5 (1 = *never*, 3 = *sometimes*, 5 = *always*). A subset of these preliminary items was sent back to three of the original scholars, who were asked for their feedback.

Pre-Testing for Content & Face Validity. Following the evaluation of a subset by expert judges, all of the items were administered to a target population for evaluation and pre-testing. Three focus groups, containing a total sample of 10 graduate students, were asked to take the draft survey and then participate in a 45-minute semi-structured discussion during which they were asked what they thought the questions meant; if the vocabulary used in the items was clear; and if there were any differing interpretations of the questions. The focus groups were recorded on Zoom, transcribed using

Otter.ai, and coded for themes and insights. The findings were then used to further refine the items for better face validity.

Quantitatively Testing for Construct Validity and Reliability. Factor analysis was chosen as the method to further refine the scale and reduce data to find a more parsimonious, useful, and applicable scale. Factor analysis is often used in scale development for such purposes, as it is a statistical method which explores the relationships between a large number of observed variables and explains those relationships using a smaller set of unobserved or latent variables (Flora & Flake, 2017).

The refined 38-item scale (Appendix A) was administered to a sample of 250 participants using two online platforms: Connect and Prolific. Based on Patton (2018) and Buckley and colleagues' (2015) call to build capacity to think evaluatively in as many people as possible, evaluators and non-evaluators alike, and given the idea that ET is considered essential not just for evaluations but for organizational development and democratic engagement, the CETS was intended to measure ET in non-evaluators as much as evaluators. As such, the sample was sourced from the online platforms Prolific and Cloud Research with very broad inclusion criteria. Both Connect and Prolific are online research platforms that recruit and manage participants for researchers by advertising a study, screening potential participants according to the researcher's inclusion and exclusion criteria, directing qualifying participants to the existing survey software (in this case, Qualtrics), and managing systems to appropriately compensate participants. To be included in this study, participants had to be over 18 years old, be fluent in English, and have at least 1 year of work experience and a high school degree. Each participant was offered the equivalent of around \$9/hr to participate in this 15–20-minute study. See Table 2 for sample demographics.

Table 2. Demographics of the Factor Analysis Sample

Characteristics	Category	Frequency	Percentage
Sex	Male	144	41.26%
	Female	101	28.94%
Age	18–28	78	31.84%
	29–39	89	36.33%
	40–50	54	22.04%
	51–61	20	8.16%
	62–72	4	1.63%
Education	High school diploma	32	13.01%
	Some college, but no degree	39	15.85%
	Associate or technical degree	24	9.76%
	Bachelor's degree	110	44.72%
	Graduate or professional degree	41	16.67%
Race	White	168	74.01%
	Black or African American	51	22.47%
	American Indian or Alaska Native	1	0.44%
	Asian	0	0.00%
	Native Hawaiian or Pacific Islander	1	0.44%
	Other	6	2.64%

Results

Research Question 1: What Dimensions Demonstrate ET?

Thematic content analysis of the literature uncovered five dimensions of ET that grounded the development of the interview protocol. According to these five dimensions, someone who thinks evaluatively considers context, engages in reflective practice, seeks out and integrates multiple perspectives, constructs sound arguments, and is engaged in a process of coming to a value judgment. These five dimensions, in addition to other questions—such as “What is the difference between critical and evaluative thinking?” and “What is the difference between evaluative thinking and practice?”—were probed further in the interviews with the nine scholars.

The thematic content analysis conducted following the qualitative coding of the transcripts from the expert interviews expanded the themes

from the literature from five to nine overlapping dimensions of ET. According to this analysis, someone who thinks evaluatively [is] reflective, is intentional/planful, forms questions of consequence, seeks out *pertinent* evidence, checks assumptions, takes a systematic approach, considers context, seeks multiple perspectives, and has an end goal of informing decisions or value judgments (often for social betterment). This last point, about the end goal of ET, emerged from predominant interviewee responses that what distinguishes evaluative thinking from critical and other ways of thinking, is the “valuing” or “value-judgment” step. Further descriptions of these dimensions can be found in Table 1.

Research Question 2: Do the Items Probing these ET Constructs Have Validity?

A 40-item scale, with at least four items probing each of these nine dimensions of ET, was administered to 10 graduate students from across disciplines at the start of three different focus

groups. After taking the survey, participants participated in a discussion about the interpretability of the items. The insights gleaned from these focus groups highlighted multiple items that were double-barreled and long: how the questions invited social desirability (i.e., participants wanting to choose *sometimes* to *always* more often from the response set); and the use of technical evaluation language which was not interpreted as it was intended. Most importantly, to this last point, for the non-evaluators in the focus group, the use of the term “value judgment” or arriving at a “value judgment” was not interpreted as an assessment of value or worth of the subject of an evaluation, as intended, but was, instead, interpreted as a judgment based on someone’s personal values.

The original scale was then refined, based on focus group participant feedback and suggestions, to improve interpretability of the commonly misinterpreted terms for those outside of the field of evaluation; the aesthetic of the survey was upgraded; the Likert scale was changed from a five- to a seven-point scale (1 = *never*, 4 = *sometimes*, 7 = *always*) to improve variance and minimize error derived by social desirability; and the cognitive recall prompt was tightened to foster participant focus on recalling consequential decision-making processes rather than on “coming

to a value judgment.” A 38-item scale resulted, for which an example item from each construct is shown in Table 3.

An exploratory factor analysis (EFA) conducted in R was used to explore whether one or more latent factors could account for dimensions of ET, and to determine if the resulting 38-item scale, probing the nine overlapping factors discovered in Stage 1 of the study, could be further refined and a more parsimonious, reliable, and valid scale could emerge (Hayton et al., 2004). The 38-question survey was administered to over 250 people. Of those, 250 participants filled out the survey completely, passing two attention checks and with no missing data, thus making up the sample for this factor analysis. Given the sample size was 250 and the ratio of respondents to items is greater than 5:1, this sample size was adequate for conducting an EFA (Gorsuch, 1983). Descriptive statistics were generated, and skew and kurtosis values were under 3 and 10, respectively, indicating normality. A Heinz-Zirkler test confirmed that the data were not multivariate normal, $HZ = 1.10$, $p < .001$. Additionally, the data were checked for multivariate outliers and 21 were detected and removed. See Appendix B for the frequencies for the original 38 items.

Table 3. Nine Dimensions of Evaluative Thinking, Descriptions, and Example Items to Probe Them

Construct	Description	Example item
Reflective	Recognizes their own biases and values and how they might impact their thinking, processes, and/or perspectives. They take the time to reflect upon their processes and consider why they choose to do things the way they do.	I analyze areas of consistency and inconsistency in my thinking.
Checks assumptions	Reflects on own assumptions, questions others' assumptions, and considers how each might influence the process or conclusion.	I ask myself how I know what I think I know.
Takes a systematic approach	Ensures their approach to data collection and analysis is consistent and can be explained, justified and defended.	I consider whether my process of arriving at a conclusion can be justified.
Forms questions of consequence	Forms questions about the evaluand that require careful consideration and thought, and the answers to those questions most likely have consequences for the evaluand. It is to acquire those answers that the thinker is developing a process.	I pose questions about the project that require careful consideration.
Intentional/planful	One who thinks through, in advance, how they will get trustworthy, reliable answers to their questions of consequence.	I think through how to get trustworthy answers to my questions about the program.
Seeks out pertinent evidence	Thinks through, in advance, who the stakeholders are, what will be considered "trustworthy" evidence to them, and how to gather this evidence.	I consider what evidence will be most valuable in justifying the conclusion.
Considers context	Seeks to understand the factors acting on the evaluand and what influence the evaluand has on its operating environment.	I seek to understand the environment in which the program is operating.
Seeks multiple perspectives	Both values and seeks out multiple perspectives on the questions of consequence.	I consider how to create avenues to hear multiple perspectives on the project.
End goal (seeks to inform decisions/social betterment):	Has an end goal of informing decision-making processes that will ultimately lead to social betterment.	When a program decision is needed, I strive to create a process that ensures a confident assessment of worth.

Note: "Evaluand" refers to the subject of an evaluation.

The data proved appropriate to undergo a factor analysis based on four criteria: the correlation between variables, the determinant, Bartlett's test of sphericity, and the KMO (Howard, 2016). A glance at the correlation table (Polivy,

2024) showed most correlations between observed variables to be greater than .30. The determinant, which assesses cases of extreme multicollinearity, was just barely greater than 0, indicating multicollinearity is not an issue. Bartlett's test of

sphericity returned a $\chi^2 = 6290.118$, $p < .001$, indicating the correlation matrix is not an identity matrix. And the KMO is 0.93, which is greater than the recommended cut off 0.70 (Howard, 2016). Together, these tests indicate that the data has enough covariance to warrant an EFA (Howard, 2016).

Next, the items were subjected to a common factor analysis to determine how many factors to retain initially. Due to the data not being multivariate normal, we used ordinary least squares to determine how many factors to retain (Howard, 2016). A general approach to factor analysis is for researchers to retain factors if they are contributing a considerable amount of variance and factors that contribute less variance should not be retained in the final model (Hayton et al., 2004). Several proven methods can be used to make factor retention decisions. A scree plot is a visual representation of the amount of variance contributed by each factor; however, for more complex models with large numbers of potential factors, it can be ambiguous and subjective (Hayton et al., 2004). A parallel analysis has been shown by Hayton and colleagues (2004) to be the most accurate method for factor retention. Therefore, since both methods are considered good ways to determine factor retention and our set of possible factors was under 10, a parallel analysis and a scree plot were both used to determine that six factors be retained *to start*.

A six-factor model using the ordinary least squares method and an oblimin rotation (due to high correlations amongst variables (Howard,

2016)² was run four times in search of the most parsimonious and representative model in which all factors had high eigenvalues, above 1; variables representing multiple factors and any variables that load onto factors at 0.40 or less were removed (Howard, 2016). The final six-factor pattern matrix can be found in Table 4.

Each factor was then named based on the common conceptual underpinnings of its items as discovered from Stage 1. Factor 1 is mostly comprised of items originally created to probe the dimension of *considers context* and thus is named as such. Factors 2, 3, and 6 retained their originally conceptualized names because the items that saliently loaded on each consisted of the original items created to probe each dimension. Factor 2 was named *seeks multiple perspectives*; Factor 3 is *intentional/planful*; and Factor 6 is called *engaged in reflective practice*.

The items which saliently load on Factor 5 are a combination of two of the nine originally conceived factors: “seeking pertinent evidence” and “systematic approach.” Items from these two dimensions combining to one factor makes sense because both are about crafting a sound, defensible argument, which is one backed up with trustworthy and reliable evidence. Therefore, Factor 5’s name returns to one of the earliest dimensions discovered in the literature review which is *constructs sound arguments*. The final factor model, with the items and factor names, can be found in Figure 1.

² Correlation table of the final six-factor model can be found in Appendix C

Table 4. Factor Pattern Matrix

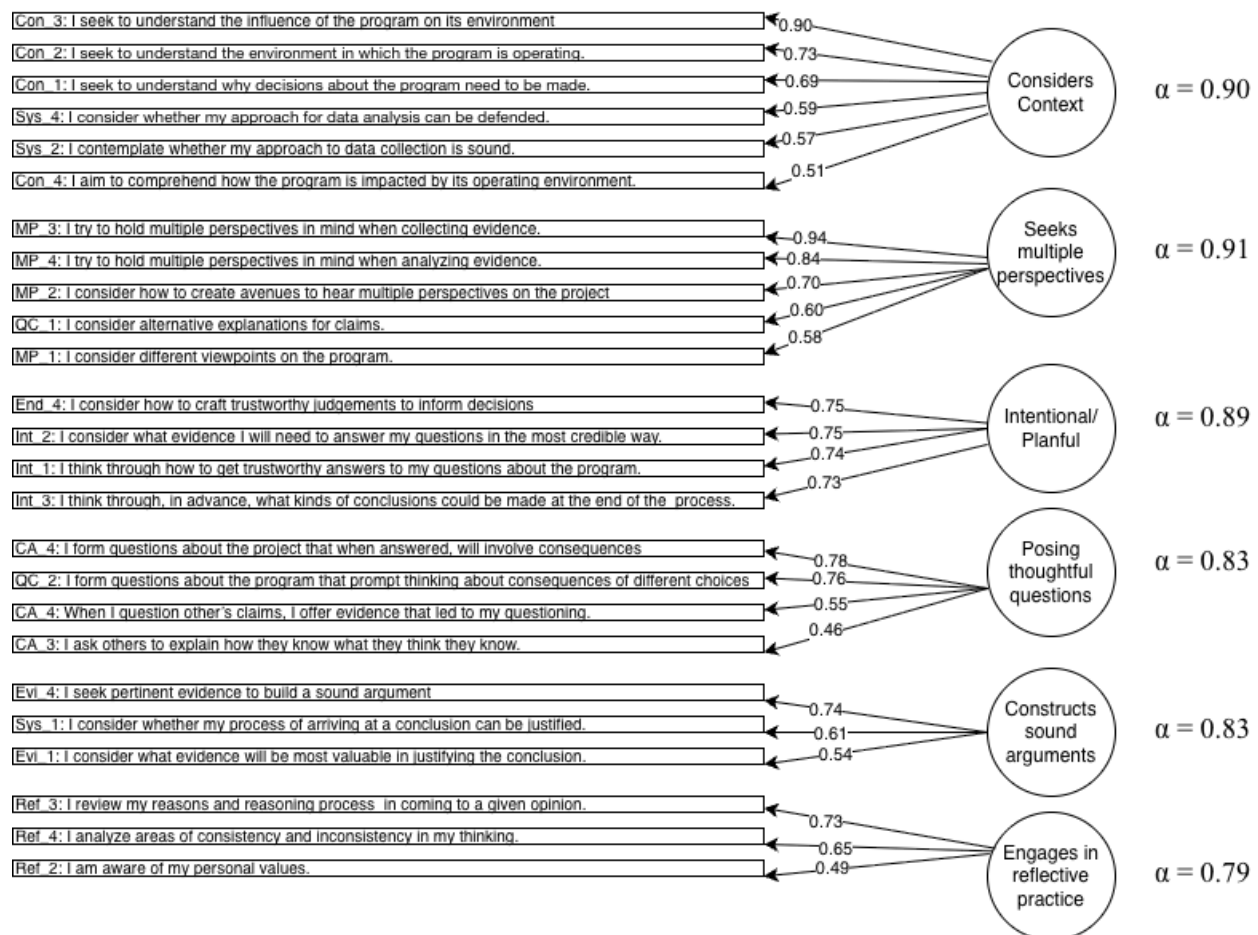
Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Communality	Unique variance
Con_3	0.90						0.73	0.27
Con_2	0.73						0.68	0.32
Con_1	0.69						0.63	0.37
Sys_4	0.59						0.60	0.40
Sys_2	0.57						0.58	0.42
Con_4	0.51						0.55	0.45
MP_3		0.94					0.85	0.15
MP_4		0.84					0.69	0.31
MP_2		0.70					0.68	0.32
QC_1		0.60					0.65	0.35
MP_1		0.58					0.64	0.36
End_4			0.75				0.75	0.25
Int_2			0.75				0.71	0.29
Int_1			0.74				0.66	0.34
Int_3			0.73				0.67	0.33
CA_4				0.78			0.62	0.38
QC_2				0.76			0.76	0.24
CA_4				0.55			0.57	0.43
CA_3				0.46			0.44	0.56
Evi_4					0.74		0.75	0.25
Sys_1					0.61		0.70	0.30
Evi_1					0.54		0.61	0.39
Ref_3						0.73	0.65	0.35
Ref_4						0.65	0.62	0.38
Ref_2						0.49	0.46	0.54

Research Question 3: Is There Internal Consistency Between These Items?

To measure the internal consistency of these scales, the Cronbach's Alpha for both the survey overall as well as the individual factors was determined. Cronbach's Alpha is a way to measure the internal consistency, or reliability, of a set of items in a scale (Crano et al., 2015). An acceptable Cronbach's Alpha coefficient is 0.75 and can range from 0.00 to

1.00, with a higher coefficient identifying more highly interrelated items and thus a more reliable set of items measuring the same construct (Crano et al., 2015). The Cronbach's Alpha for the survey was $\alpha = 0.95$. The Cronbach's Alpha for each of the factors was also determined, and the results can be found in Figure 1. All the Cronbach's Alpha coefficients are greater than $\alpha = 0.75$, indicating this scale has good internal consistency.

Figure 1. Final Factor Model for the Claremont Evaluative Thinking Scale (CETS) after EFA



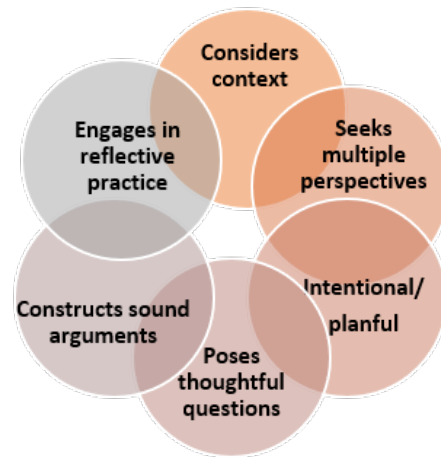
Discussion

This study looked at the construct of ET and developed a scale named the Claremont Evaluative Thinking Scale (CETS). The scale has 25 items, which are internally consistent and have face validity and which operationalize six dimensions of ET, thus enhancing our ability to measure this construct. Scholars have written and discussed the concept of ET and its added value to successful evaluations. However, to date, no reliable and valid way to operationalize and measure the concept has been widely used to test these claims. This revised six-factor scale strengthens researchers' capacity to do just that. For example, Vo and colleagues (2018) share the idea that "ET reflects the ability to creatively arrive at evaluation-specific solutions by determining what combinations of methods, tools, techniques, and theories are appropriate *in light of contextual particulars*" (p. 107). The combination of Factor 1, *considers context*, and Factor 3,

intentional/planful, on the revised evaluative thinking scale probes this concept and enables measurement of how often individuals are considering context and intentionally thinking about what data collection strategies are appropriate given this context. Additionally, Factor 5, *constructs sound arguments*, operationalizes part of Vo's (2013) definition of ET, which states, "It is the process by which one marshals evaluative data and evidence *to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion*" (p. 107). And Factors 2 (*seeks multiple perspectives*), 3 (*poses thoughtful questions*), and 6 (*engages in reflective practice*) operationalize the piece of Buckley and colleagues' (2015) definition of ET, which states "pursuing deeper understanding through reflection and perspective taking" (p. 378). The CETS offers a positive step toward operationalizing certain theoretical definitions of ET and offers us a chance to begin measuring individual-level aspects of the construct and create a body of empirical evidence to

test the theories about ET posited to date, such as its added value to evaluations and society.

Figure 2. Six Dimensions of Evaluative Thinking Discovered Through the EFA



The dimensions probed by the CETS are both similar to and different from those the ETI deemed important and sought to operationalize. Similarly to the ETI, the CETS probes the ideas of forming thoughtful questions, being reflective, and seeking alternatives by seeking out multiple perspectives. These concepts are all represented in the factor *poses thoughtful questions and seeking alternatives*, which is one of the three factors discovered in McIntosh and colleagues' (2020) EFA of the ETI. However, the other two factors of the ETI discovered through this EFA, "describing and illustrating thinking" and "expressing belief in the value of evaluation," did not surface in the Stage 1 findings of this study and thus were not tested nor included in this revised scale. Furthermore, the idea of ET occurring in an evaluation context did not emerge in this study, although it is a very present concept in Buckley and colleagues' (2015) definition of ET. Likewise, the idea of being intentional and planful did not emerge as a dimension in the ETI or in any of the leading definitions of ET but arose often in Stage 1 of this study. The idea of being intentional, from the onset and throughout the process, about *how* to engage in the process, is a dimension that warrants continued exploration. With this revised scale, researchers

can now probe this dimension further, as well as any predictive qualities it may have.

While the six-factor model of ET represented in the data is a more parsimonious model than the nine factors which emerged from Stage 1 of this study, one important idea, that of the "end goal," did not surface in the final model. One of the earliest dimensions of the construct, which surfaced from the literature review, was that someone who is thinking evaluatively is "engaged in a process of *coming to a value judgment*." This idea signaled the end goal of ET is to arrive at a value judgment. This idea was probed further in the interview stage by asking, "What is the end goal of evaluative thinking?" The responses from the experts varied from "to arrive at a value judgment" to "to inform decision making" to "social betterment." Furthermore, one of the key findings from Stage 1 of the study is that the value judgment "step" is what sets ET apart from critical thinking. Quotes from the interviews illustrate this point:

However, what I see as the difference is ET also incorporates other aspects or other elements that may not always be present in critical thinking, namely, the valuing piece, the rational process of ascribing merit worth or significance. You can think critically, but you

may not derive a value judgment at the end (Interviewee A, 2022).

The items we created to probe this dimension of “the end goal of evaluative thinking” did not hang together, nor did they load saliently onto any other factor. This most likely occurred because the created items were probing the different possible end goals (i.e., arriving at a value judgment, informing decisions, social betterment) rather than probing one of those end goals, which, in this case, might have more aptly been “arriving at a value judgment.” Another consideration for why this factor did not emerge is that it might have been captured in the prompts throughout the survey rather than by the items themselves. Participants were regularly reminded to consider themselves in the process of informing a decision and to rate how often they consider each item when in that decision-making process. Therefore, the items regarding “the end goal of evaluative thinking” might have been confusing given they were already imagining themselves in a decision-making process.

The failure of valuing-related items to cohere statistically is itself an empirical finding, not merely a limitation. This finding is consistent with the theoretical position that valuing in evaluative thinking is dialogic, situational, and often negotiated, making it poorly suited to static self-report frequency measures. We see the present findings as complementary to theorists’ arguments that valuing emerges through interaction, deliberation, and practical judgment (Gates, 2018; Schwandt, 2018), and we view observational, dialogic, and case-based methods as essential complements to the present instrument.

Strengths and Limitations

This study had both strengths and limitations. One of the key strengths of this study was the thorough process of conducting both a literature review and interviews with experts on the topic. Conducting interviews after reviewing the literature allowed us to develop a greater understanding of ideas in the literature that were previously ambiguous or in conflict. Furthermore, as we interviewed experts, more current thinking on earlier published ideas emerged and was included in this attempt to define dimensions of ET—for example, the idea of being “intentional/planful” emerged from the interviews and not the literature, as did an understanding of how ET can exist outside of evaluation contexts. This two-step qualitative process allowed for a more expansive and current understanding of ET.

A limitation of this study was the random respondent sample collected by Connect and Prolific. Collecting data from people on these platforms was limiting in two ways. The first limitation was in trying to create language and cognitive recall prompts that were generic enough that a random sample comprised of those without evaluation experience would be able to understand what was being asked and answer the questions well. While the focus groups allowed for troubleshooting the language and ensuring the intended meaning of each item was understood, the fact that the idea of “arriving at a value judgment” did not resonate as it was supposed to meant that researchers were limited in their ability to probe this dimension sufficiently. And as was acknowledged earlier, this is an important dimension to include in any attempt to measure ET. The second limitation with this sample is that a clear group did not emerge to whom we could generalize the findings.

Another limitation is that more might now be known about ET and the types of items that might probe the construct and its dimensions; however, from the limited scope of this study, it is not yet known how this scale correlates with other ET instruments (convergent validity), nor how it associates or doesn’t associate with other scales or constructs (divergent validity). However, this study did create a refined, reliable scale which can now, in a future study, undergo tests of such validity, especially of those dimensions that were uncovered.

A final limitation of this study is the notion of using solely a self-report instrument to attempt to measure a concept as complex and dynamic as ET. As noted in the introduction of this paper, ET is a contextual, relational, value-laden construct which, by definition, is difficult to measure via self-report and cross-sectionally. While the exercise of developing a valid and reliable self-report scale is an important contribution toward concretizing understanding of the concept and even measuring aspects of it, the self-report tool could only benefit from being used in conjunction with other innovative measures that can capture some of the concepts not included in this scale, such as more objective measures, as discussed below. At the same time, we are often limited in our resources and ability to use other methods, and self-report is what we have. While self-report instruments are limited, the CETS offers an accessible instrument and has laid the groundwork upon which such innovative complementary measures can be built and utilized together.

Areas for Future Research

One area for future research is to develop items to probe that missing dimension of the end goal of ET and to test for divergent validity with this scale and a critical thinking scale. A second area for future research is to test the CETS on a new sample of a similar population and conduct a confirmatory factor analysis (CFA) on this data to see if the model holds. A third area for future research would be to revise or simplify the items probing the idea of the end goal of ET to better reflect the idea of “coming to a value judgment” and test the scale again to probe that dimension more thoroughly. As the scale continues to be refined through adding additional factors and conducting a CFA, another area of research would be to administer this scale alongside other valid measures of belonging, intra-team communication, and purpose to understand the predictive qualities of ET as well as its added value. Once the predictive qualities of ET are better understood, the scale could then be used by researchers, evaluators, or organizations to better understand in which areas or dimensions of ET their colleagues are strong or weak and design interventions to strengthen or enhance those dimensions of ET amongst their team. A final yet possibly most important area of research could be to use the dimensions of ET discovered in this study, especially those which did not emerge as factors on the CETS, to develop other complementary ways to measure ET to get beyond the noted challenge of self-report surveys being subject to bias and social desirability (Wingate et al., 2022). Complementary measurement approaches could illuminate dimensions of ET that are beyond the individual level and are less probe able with a self-report such as the relational and contextual dimensions discussed throughout the introduction.

Conclusion

ET is a much-discussed topic in the field of evaluation, yet it is a challenging concept to understand and differentiate from other ways of thinking and reasoning. At the same time, theories of the value of ET in both evaluation and non-evaluation contexts abound. There are theories about how thinking evaluatively in evaluation contexts can lead to more smooth and successful evaluations; in organizational settings it could offer a shared way of thinking that promotes performance; and in larger societal contexts it can lead to strengthened democracies. However, we cannot know if ET has any impact on the

aforementioned or on other aspects of society without being able to measure it. This study offers a first step toward more cumulative measurement work of ET by empirically testing whether a subset of widely shared ET dimensions can be operationalized in a reliable way. The six dimensions of ET probed with this reliable and valid scale are *considers context, seeks multiple perspectives, intentional/planful, poses thoughtful questions, constructs sound arguments, and engages in reflective practice*. This initial scale-development study established psychometric feasibility, though not conceptual exhaustiveness and sets the stage for a longer program of validation, refinement, and triangulation. Furthermore, this exploratory study lays the groundwork to develop additional innovative measures for ET and thus more deeply understand its impact.

References

- Archibald, T., Sharrock, G., Buckley, J., & Cook, N. (2016). Assumptions, conjectures, and other miracles: The application of evaluative thinking to theory of change models in community development. *Evaluation and Program Planning*, 59, 119–127. <https://doi.org/10.1016/j.evalprogplan.2016.05.015>
- Buckley, J. & Archibald, T. (2011). The Evaluative Thinking Inventory. Cornell University.
- 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.
- Chauveron, L. M., Urban, J. B., Samtani, S., Cox, M., Moorman, L., Hargraves, M., Buckley, J., & Linver, M. (2021). Promoting evaluation in youth character development through enhanced evaluation capacity building: Empirical findings from the PACE Project. *New Directions for Evaluation*, 169, 79–95. <https://doi.org/10.1002/ev.20447>
- Crano, W. D., Brewer, M. B., & Lac, A. (2015). *Principles and methods of social research* (3rd ed). Routledge. <http://site.ebrary.com/id/10921638>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third edition). SAGE.
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation.

- Canadian Journal of Behavioural Science*, 49(2), 78–88.
<http://dx.doi.org/10.1037/cbs0000069>
- Gates, E. F. (2018). Toward valuing with critical systems heuristics. *American Journal of Evaluation*, 39(2), 201–220.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). L. Erlbaum Associates.
- Hargraves, M., Buckley, J., Urban, J. B., Linver, M. R., Chauveron, L. M., Samtani, S., Moorman, L. (2021). Resonance, stickiness, and the value propositions of Evaluation Capacity Building: Key takeaways and future directions. *New Directions for Evaluation*, 2021(169), 97–116
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205.
<https://doi.org/10.1177/1094428104263675>
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What are we doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51–62.
<http://dx.doi.org/10.1080/10447318.2015.1087664>
- Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports*, 6(1), 40–41.
- Leviton, L. C. (2014). Some underexamined aspects of evaluation capacity building. *American Journal of Evaluation*, 35(1), 90–94.
<https://doi.org/10.1177/1098214013502844>
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). SAGE Publications.
- McIntosh, J. S., Buckley, J., & Archibald, T. (2020). Refining and measuring the construct of evaluative thinking: An exploratory factor analysis of the Evaluative Thinking Inventory. *Journal of MultiDisciplinary Evaluation*, 16, 104–117.
- Papadas, K.-K., Avlonitis, G. J., & Carrigan, M. (2017). Green marketing orientation: Conceptualization, scale development and validation. *Journal of Business Research*, 80, 236–246.
<https://doi.org/10.1016/j.jbusres.2017.05.024>
- Patton, M. P. (2022, February 28). *Research Focused Approaches to Evaluation* [Lecture]. Comparative evaluation theories, Claremont Graduate University.
- Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. *New Directions for Evaluation*, 158, 11–28.
<https://doi.org/10.1002/ev.20325>
- Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making: Evaluative thinking as a collaborative social practice. *New Directions for Evaluation*, 158, 125–137.
<https://doi.org/10.1002/ev.20318>
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Sage Publications.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a Definition of the ECB Process: A Conversation with the ECB Literature. *New Directions for Evaluation*, 93, 1–25.
- Symonette, H. (2004). Walking pathways toward becoming a culturally competent evaluator: Boundaries, borderlands, and border crossings. *New Directions for Evaluation*, 102, 95–109.
<https://doi.org/10.1002/ev.118>
- Tan, L., Kocsis, A., & Burry, J. (2023). Advancing Donald Schön's Reflective Practitioner: Where to Next? *Design Issues*, 39(3), 3–18.
https://doi.org/10.1162/desi_a_00722
- The Bruner Foundation. (2014–2015). *Effectiveness initiatives: Evaluative thinking*. <http://www.evaluativethinking.org/evalthink.html>
- Tovey, T. L. S. & Archibald, T. (2023). The relationship between reflective practice, evaluative thinking, and practical wisdom. In M. Hurteau & T. Archibald (Eds.), *Practical Wisdom for an Ethical Evaluation Practice* (pp. 87–101). Information Age Publishing.
- Vo, A. T. (2013). Toward a definition of Evaluative Thinking [UCLA].
<https://escholarship.org/uc/item/26t5x5f6>
- Vo, A.T. and Archibald, T. (2018), *Editors' Notes*. *Evaluation*, 2018: 7–9. <https://doi.org.ccl.idm.oclc.org/10.1002/ev.20321>
- Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of Evaluative Thinking. *New Directions for Evaluation*, 158, 29–47.
<https://doi.org/10.1002/ev.20324>
- Wingate, L. A., Robertson, K., FitzGerald, M., Rucks, L., Tsuzaki, T., Clasen, C., & Schwob, J. (2022). Thinking outside the self-report: Using evaluation plans to assess evaluation capacity building. *American Journal of Evaluation*, 43(4), 515–538.
<https://doi.org/10.1177/10982140211062884>

Appendix A: Original 38-Item Scale

Original 38 Items		
DIMENSION	ITEM	Abbreviated item
REFLECTIVE:	I recognize my personal biases.	Ref_1
	I am aware of my personal values.	Ref_2
	I review my reasons and reasoning process in coming to a given opinion.	Ref_3
	I analyze areas of consistency and inconsistency in my thinking.	Ref_4
CHECKS ASSUMPTIONS:	I appreciate it when others (e.g., people with vested interest in the decision being made) ask me to explain how I got to my claim.	CA_1
	I ask myself how I know what I think I know.	CA_2
	I ask others to explain how they know what they think they know.	CA_3
	When I question others' claims, I offer evidence that led to my questioning.	CA_4
SYSTEMATIC APPROACH	I consider whether my process of arriving at a conclusion can be justified.	Sys_1
	I contemplate whether my approach to data collection is sound.	Sys_2
	I think about making reliable claims to inform decision-making.	Sys_3
	I consider whether the approach I choose for data analysis can be defended.	Sys_4
FORMS QUESTIONS OF CONSEQUENCE:	I consider alternative explanations for claims.	QC_1
	I form questions about the project that, when answered, will involve consequences.	QC_2
	I form questions about the program that prompt thinking about the consequences of different choices.	QC_3
	I pose questions about the project that require careful consideration.	QC_4
INTENTIONAL/PLANFUL:	I think through how to get trustworthy answers to my questions about the program.	Int_1
	I consider what evidence I will need to answer my questions in the most credible way.	Int_2
	I think through, in advance, what kinds of conclusions could be made at the end of the assessment process.	Int_3
	I identify consequences of actions which might result from my conclusions.	Int_4
ACTIVELY SEEKS OUT PERTINENT EVIDENCE:	I consider what evidence will be most valuable in justifying the conclusion.	Evi_1
	I consider what evidence will be most respected by others who have a stake in the decision being made.	Evi_2
	I consider what evidence will be most respected by the decision makers.	Evi_3
	I seek pertinent evidence to build a sound argument.	Evi_4
CONSIDERS CONTEXT	I seek to understand why decisions about the program need to be made.	Con_1
	I seek to understand the environment in which the project is operating.	Con_2

		I seek to understand the influence of the program on its environment (physical & people).	Con_3
		I examine the interrelationships among people, policies, and processes.	Con_4
		I aim to comprehend how the program is impacted by its operating environment.	Con_5
SEEKS PERSPECTIVES	MULTIPLE	I consider different viewpoints on the program.	MP_1
		I consider how to create avenues to hear multiple perspectives on the project.	MP_2
		I try to hold multiple perspectives in mind when collecting evidence.	MP_3
		I try to hold multiple perspectives in mind when analyzing evidence.	MP_4
END GOAL		When a program decision is needed, I strive to create a process that ensures a confident assessment of worth.	End_1
		I craft processes that help assess value in order to inform decisions that lead to action.	End_2
		When I craft processes to inform a programmatic decision, I consider how it will improve society.	End_3
		I consider how to craft trustworthy judgements to inform decisions.	End_4
		When making a judgement about the value of something, I consider whether it can be justified.	End_5

Appendix B: Likert Frequencies of the 38 Items

Item	1	2	3	4	5	6	7
Ref_1	2	5	13	69	58	71	29
Ref_2	1	0	7	15	55	86	83
Ref_3	2	3	6	24	65	87	60
Ref_4	2	2	19	50	70	61	43
CA_1	2	8	13	31	60	83	50
CA_2	5	16	28	65	47	60	26
CA_3	2	15	22	58	72	47	31
CA_4	3	4	15	41	71	62	51
QC_1	9	21	21	70	56	49	21
QC_2	6	7	22	59	72	57	24
Sys_3	1	1	11	30	79	70	55
QC_4	0	4	18	48	73	71	33
Int_1	1	7	17	52	60	69	41
Int_2	1	5	14	28	61	88	50
End_4	1	2	7	33	74	73	57
Int_3	1	4	6	51	59	75	51
Int_4	1	2	8	35	73	79	49
Evi_1	0	1	6	39	59	77	65
Evi_3	4	4	13	41	70	73	42
Evi_4	0	0	1	25	54	80	87
Sys_1	1	0	8	31	55	86	66
Sys_2	1	4	12	35	64	73	58
Sys_4	2	4	12	41	50	80	58
Con_1	1	2	16	34	58	80	56
Con_2	1	4	9	32	76	68	57
Con_3	2	4	12	26	65	73	65
Con_4	1	5	19	48	69	63	42
Con_5	1	12	10	43	63	78	40
MP_1	1	2	7	21	63	75	78
MP_2	1	6	12	44	70	59	55
MP_3	1	4	14	40	55	76	57
QC_1	1	1	11	38	53	84	59
MP_4	3	3	6	47	60	72	56
End_1	1	3	4	34	74	70	61
End_2	1	7	8	44	76	69	42
End_3	8	14	22	47	53	59	44
End_5	0	3	8	24	68	88	56
Evi_2	2	6	15	42	57	72	53

Appendix C: Correlation Table of Final Items in a Six-Factor Model

[illegible]

[illegible]