

# Validating an Evaluation Checklist Using a Mixed Method Design<sup>1</sup>

*Wes Martz*

A checklist is a mnemonic device that consists of a list of activities, items, and criteria used to perform a certain task. When used in evaluation, checklists provide guidance for the collection of relevant evidence used to determine the merit, worth, or significance of an evaluand. The inherently systematic process found in the use of a checklist makes it highly relevant and useful for evaluative purposes. As described by Scriven (2005), checklists reduce the chance of forgetting to check something important, are easier for the layperson to understand and validate than most theories or statistical analyses, and reduce the influence of the halo effect by forcing the evaluator to consider each relevant dimension of merit. In addition, checklists incorporate a vast amount of specific knowledge about the particular evaluand in a parsimonious manner and facilitate the evaluation task. Consequently, checklists can contribute to the improvement of validity, reliability, and credibility of an evaluation (Scriven, 2007).

The value of checklists for evaluation, as summarized above and written about extensively (e.g., Hales & Pronovost, 2006; Keil, Li, Matthiassen, & Zheng, 2008; Persaud, 2007; Sanders, 2001; Schroeter, 2008; Scriven, 2005, 2007; Stufflebeam, 2000, 2001; Stufflebeam & Shinkfield, 2007), is generally accepted. There are, however, potential negative impacts that could result from poorly designed or misapplied checklists, including checklist fatigue, decreased reliability by adding unnecessary complexity to the evaluation process, and delays in the completion of the evaluation resulting from burdensome processes. In addition, what could be referred to as evaluation myopia—the inability of the evaluator to identify side effects or side impacts due to rigid application and non-reflective use of the checklist—may also

---

<sup>1</sup> Martz, W. (2010). Validating an evaluation checklist using a mixed method design. *Evaluation and Program Planning*, 33, 215-222. Available at <http://dx.doi.org/10.1016/j.evalprogplan.2009.10.005>.

develop. Despite the potential hazards that could result from checklist use, a literature review conducted across fields and industries comprising 178 peer-reviewed articles, university affiliated publications, and government published documentation on checklists could find no published data indicating adverse or negative effects of checklist use (Hales, Terblanche, Fowler, & Sibbald, 2008). This lack of published documentation, however, does not lessen the importance of validating a checklist—particularly when the checklist is intended for evaluative purposes. This article outlines one approach for validating an evaluation checklist by presenting a case example of a validation study applied to the Organizational Effectiveness Evaluation Checklist (OEC). Following a brief overview of the OEC, the methods used to validate the checklist are presented followed by a discussion of the limitations of the study.

#### Background on the OEC

The OEC was developed to advance evaluation in organizational settings, particularly business and industry (Martz, 2008). It is a tool for professional evaluators, organizational consultants, and management practitioners to use when designing, performing, or metaevaluating an evaluation of organizational effectiveness. The OEC outlines a systematic process to assess organizational effectiveness that addresses many of the shortcomings found in the most prevalent models for evaluating organizational effectiveness presented in the management and evaluation literature (e.g., Altschuld & Zheng, 1995; Owen & Lambert, 1998; Papadimitriou, 2007; Ridley & Mendoza, 1993; Schweigert, 2006). These deficiencies include, but are not limited to, narrow or inappropriate value premises, indiscriminate rationale for selecting criteria of merit, inconsistent level of analysis, overlapping and conflicting criteria of merit, limited contextual sensitivity, and a lack of ethical considerations.

Consistent with the evaluation for leadership approach suggested by Owen & Lambert (1998), the OEC frames the assessment of organizational effectiveness by considering the organization as a whole while individuals and programs are a secondary level of concern. That is, the focus of the OEC is not on a specific program or initiative of an organization; the focus is on the organization itself. From a practical perspective, the OEC is explicitly focused on the input-output cycle of the organization, not on the frame of reference of the actors of the organization under study. This shifts the attention from the value perspective of the researcher or a particular organizational member to value premises based on organizational needs, definition, ethics, and other valid and defensible sources.

The operational definition of organizational effectiveness used in the OEC is derived from the definitional premise of organization (i.e., deliberate, planned, and purposeful), its systemic role in modern society, and influences from within and outside organizational boundaries. Hence, organizational effectiveness is defined as the extent to which the organization provides sustainable value through the purposeful transformation of inputs and exchange of outputs, while minimizing harm from its actions. The term sustainable value in this definition refers to the value ascribed by a particular stakeholder. For example, owners may define value as wealth creation. Employees may define value as fair compensation for the work they provide. Customers may define value in terms of quality versus price. Communities may define value as a measure of the financial contributions made by the organization to the community chest, whether benevolent or through taxation. Therefore, this definition of organizational effectiveness addresses organizational survival and maximization of returns through the purposeful transformation of inputs, exchange of outputs, and adaptability of an organization to deal with

the external environment, all the while minimizing harm created by the actions of the organization.

The OEC is an iterative, weakly sequential checklist with twenty-nine checkpoints grouped into six common evaluation steps: (1) establish the boundaries of the evaluation; (2) conduct a performance needs assessment; (3) define the criteria of merit; (4) plan and implement the evaluation; (5) synthesize performance data with values; and (6) communicate and report evaluation findings (see Appendix A). Embedded within the framework are twelve universal criteria of merit for evaluating organizational effectiveness. Building on Scriven's (1994) seven general guidelines for a list of criteria of merit, five additional guidelines that pertain specifically to the development of the list of criteria for assessing organizational effectiveness were considered:

- The criteria should be limited to the organizational level of analysis.
- The criteria should reflect the relation between the organization and its environment.
- The criteria should allow for the uniqueness of the organization.
- The criteria should include both the means and ends of organizational activity.
- The criteria should be stable yet provide the necessary latitude for organizational change and variability over time.

The universal criteria of merit make up the core of the OEC and are a key differentiator of the approach from other organizational effectiveness evaluation models including goal-based, systems, process, strategic constituencies, and the competing values framework.<sup>2</sup> The criteria of merit included in the OEC can be grouped into four dimensions to illustrate the connection with the operational definition of organizational effectiveness stated earlier. The dimensions include

---

<sup>2</sup> See Martz (2008) for a critical review of these models.

(1) purposeful (efficiency, productivity, and stability); (2) adaptable (innovation, growth, and evaluative); (3) sustainable (fiscal health, output quality, information management, and conflict and cohesion); and (4) ethical (intra-organizational and extra-organizational harm minimization). The universal criteria of merit capture both the means and ends of organizational activities, allow for the comparative study of organizations or subunits, and recognize the uniqueness of the organization being evaluated. When combined with the contextual criteria of merit identified from the performance needs assessment, the list provides an inventory of the characteristics and properties of an effective organization.

Although the basic framework of the OEC is intended be applied to nearly all organizations, the specific measures (i.e., performance indicators) used may differ according to the organization type (e.g., for-profit or nonprofit), purpose, or other contextual matters. In addition, the inclusion of universal criteria of merit and contextual criteria of merit provide the user with the ability to be situation-specific while working within a general paradigm of organizational effectiveness. This facilitates comparisons across studies, despite potentially varying indicators or measures from organization to organization.

#### Overview of the Validity Study Design

Two separate investigations were conducted to empirically validate the OEC. The first investigation was a nonexperimental study that sought input from subject matter experts (e.g., organizational and evaluation theorists) and targeted users (e.g., business consultants, managers, and professional evaluators) regarding the merit of the OEC. The second investigation applied the OEC to a real-world evaluation to assess the effectiveness of an organization. The methods applied to this study included qualitative and quantitative survey research and case study research. The use of a mixed method approach was deemed important to achieve the primary

benefits of *development* and *expansion* of the checklist (Greene, Caracelli, & Graham, 1989). The use of an expert panel to provide critical feedback served to facilitate the development and refinement of the OEC prior to application in the second investigation. In addition, the expert panel was considered an integral design element to demonstrate content relevance and representativeness along with the substantive and consequential aspects of validity. The case study research method facilitated learning about the practicality of the OEC in its contextual setting thereby extending the scope and breadth of the validity study (Yin, 2009).

#### *Expert Panel Critical Review*

The purposive and convenience sample was comprised of fifteen participants of which eight were female and seven were male. The sampling procedure sought participants with expertise in organizational research, evaluation, or management. What's more, the expert panel members represented a diverse array of professional and disciplinary areas, including organizational performance improvement, management, engineering, evaluation research, sociology, organizational development and training, education, change management, business consulting, project management, marketing, and accounting. All participants indicated "a great deal" or "extensive" professional experience in research (60%), evaluation (73%), or management (73%). Thirteen of the participants were of U.S. nationality, while one participant was from central Europe and another from Australia.

The instruments used in the study included the OEC, a critical feedback survey, and a sociodemographic questionnaire. All participants were asked to review the OEC for providing critical feedback, focusing on the strengths of the checklist and areas to improve, and identify items that were missing or confusing. Participants were encouraged to write directly on the checklist and asked to complete the critical feedback survey after examining the checklist.

The critical feedback survey was a pen-and-paper survey designed to collect specific data to evaluate the OEC based on Stufflebeam's (2000) eight criteria for evaluating a checklist including applicability to the full range of intended uses, clarity, comprehensiveness, concreteness, ease of use, fairness, parsimony, and pertinence to the content area. Each closed-ended item used an interval response format from 1 to 9, where 1 = strongly disagree and 9 = strongly agree. The rationale underlying the number of scale points used was that an expanded scale would allow respondents to more accurately express their positive or negative feelings about the OEC, thereby increasing the sensitivity to detect differences that may be less evident when using fewer scale points (i.e., item response options). In addition, the expert panel members' deep subject matter knowledge and homogeneous frames of reference, based on the selection criteria, supported the use of a response scale with a higher degree of discrimination (Chang, 1994). The data from the expert panel was analyzed using descriptive statistics and tests of significance for the quantitative survey items, while the qualitative data was analyzed using summative content analysis (Hsieh & Shannon, 2005).

The critical feedback survey narrative section asked the expert panel members four open-ended questions regarding the strengths and weaknesses of the checklist, items missing from the checklist, and recommendations for improving the checklist. The recording units of analysis included any character string that referred to a specific idea. In all cases, the recording unit consisted of a single word (e.g., comprehensive) or phrase (e.g., clearly outlined process). The entire written response for the particular question made up the context unit. The small sample size of expert panel participants allowed for the analysis of all written comments made on the survey form.

A total of 139 recording units were classified into one of eight categories. Of these 139 units, 71 were listed as key strengths of the checklist, 29 as key weaknesses, 20 as items missing from the checklist, and 19 as items to improve. The categories used for classification purposes included the same eight criteria for evaluating a checklist that were used in the quantitative analysis. A representative sample of recording units and their classifications are shown in Table 1.

Table 1: Representative sample of recording units and classifications

	<i>Recording unit</i>	<i>Classification</i>
<i>Strength</i>	Clearly defined steps to a rational end.	Clarity
	Brevity.	Parsimony
	Sound theory.	Concrete
<i>Weakness</i>	Based on the audience, I believe more details/examples may have been useful to illustrate particular concepts.	Clarity
	Response cost for client may be high.	Parsimony
	Checklist requires modifications for use in governmental and other not-for-profit organizations.	Applicability
<i>Missing</i>	Diversity of approaches and perspectives.	Pertinence
	Context valuables.	Applicability
	More about the concept of criteria of merit would be desirable.	Clarity
<i>Improve</i>	Beef-up and provide more thorough description of the criteria of merit.	Comprehensiveness
	More illustrations of concepts using examples.	Clarity
	While it includes some measures that might be cross-sector, that would still require some more work for evaluating nonprofits.	Applicability

In an effort to compare the responses from the closed-ended items and open-ended narrative critique, a net strength score was created for the qualitative feedback. Net strength is the difference between the normalized frequencies for the criterion's identified strengths and weaknesses. The normalized frequency is a proportional value calculated by dividing the frequency of the criterion's strengths (weaknesses) by the total number of strengths (weaknesses). The net strength score provided a ranking of the perceived strengths of the OEC based on each criterion. Table 2 presents the frequency of identified strengths and weaknesses for each construct, the normalized values, and the net strength score. The rank order of each construct based on the net strength score can be compared to the rank order of each construct based on the mean scores from the closed-ended survey items to identify similarities and differences between the quantitative and qualitative responses and to triangulate the findings across multiple measures.

Table 2: Net strength score

	<i>Strength</i>		<i>Weakness</i>		<i>Net Strength</i> ( $S' - W'$ )
	<i>Frequency</i> ( $S$ )	<i>Normalized</i> ( $S'$ )	<i>Frequency</i> ( $W$ )	<i>Normalized</i> ( $W'$ )	
Applicability	6	8	2	7	1
Clarity	13	18	7	24	-6
Comprehensiveness	11	15	3	10	5
Concreteness	10	14	2	7	7
Ease of use	11	15	5	17	-2
Fairness	7	10	3	10	0
Parsimony	6	8	5	17	-9
Pertinence	7	10	2	7	3

The net strength score is plotted against the mean score to produce the output shown in Figure 1. In this study, concreteness and pertinence criteria were ranked in the top half of both the qualitative and quantitative responses, while ease of use and parsimony criteria were ranked in the bottom half of both sets of responses. Although high mean scores for fairness and clarity placed these two criteria in the top half of the quantitative responses, they were in the lower half of the qualitative responses based on the net strength score. With respect to comprehensiveness and applicability to the full range of intended uses, both were ranked in the bottom half of the quantitative survey mean scores. However, both of these criteria were ranked in the top half of the qualitative responses based on the net strength score. This integrated perspective suggests that the OEC performed best on concreteness and pertinence to the content area, and was weakest on ease of use and parsimony.

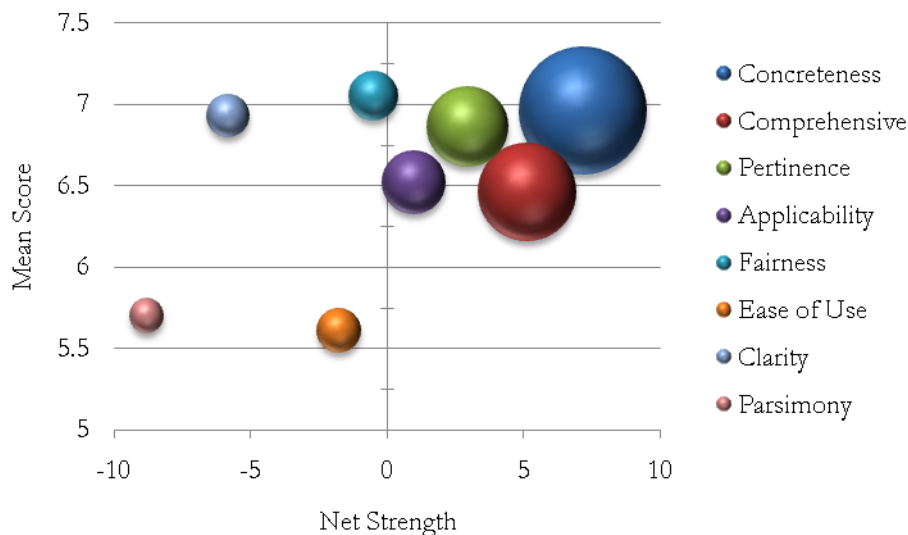


Figure 1: Net strength and mean score plot

As noted earlier, expert panel members were encouraged to make written comments directly on the OEC document. A total of 97 remarks were handwritten on the checklists; an

additional 31 were typewritten and submitted with the checklist and survey. These remarks were useful in further developing and refining the checklist to improve its ease of use. Based on the findings from phase one of the validation study, revisions were made to the checklist to build on the strengths and address deficiencies identified by the expert panel members. A keen focus was placed on identifying opportunities to modify the checklist to improve the lowest ranked evaluative criteria based on the net strength scores. The edits made to address these aspects consisted primarily of incorporating clarifying terms and explicatory comments to each checkpoint.

The most significant structural change made to the OEC was specifically intended to facilitate the efficient use of the checklist and increase its parsimony. The structural change amalgamated the distinct components of several steps and resulted in reducing the number of steps in the OEC from seven to six. It also eliminated overlapping activities that contributed to the perception of the checklist as excessively complex. As suggested by Stufflebeam (2000), consideration was given to potentially conflicting changes that could be made to the checklist to ensure the changes were mutually reinforcing and made sense for improving the checklist as a whole. As a practical matter, all suggestions regarding edits to grammar (e.g., punctuation, syntax, and semantics) were addressed as appropriate.

#### *Field Study Using the OEC*

An evaluation using the revised OEC to assess an organization's effectiveness was conducted for the second investigation. The purpose of conducting a real-world evaluation was to answer the question, "How practical is the OEC process for evaluating organizational effectiveness?" and was intended to provide evidence for consequential and substantive validity of the checklist. Each step and checkpoint outlined in the OEC was strictly followed to ensure

fidelity to the OEC process. Following the completion of the evaluation, a semi-structured interview was conducted with the evaluation client to obtain the client's perspective on the evaluation with respect to its validity, credibility, utility, cost-effectiveness, and ethicality (Coryn, 2007). In essence, this was a formative metaevaluation intended to provide a modest level of assessment that could be used to detect and correct deficiencies in the OEC process (Cooksy & Caracelli, 2005). The data from the case study was analyzed using a hypothesis-generating process (Corbin & Strauss, 2008) where the goal was to identify evidence that supported or undermined the practicality of the OEC process.

The criteria used to select the case study organization required that the organization be a stand-alone entity (i.e., not a division or program within a larger organization), in operation for at least three years, and willing to provide access to internal records and senior-level management. The participating organization was a privately held, for-profit company specializing in storytelling using a variety of print and electronic media. The U.S.-based company was founded in 2003 and had sixteen full-time employees at the time of the evaluation.

The organization's structure was moderately complex, characterized by a modest degree of horizontal differentiation. For example, employees responsible for video production were specialized with respect to their specific function and were not expected to perform tasks outside of video production. Similarly, employees responsible for writing were not expected to perform video production or other unrelated tasks. While both functions were responsible for storytelling, their functional specialization resulted in the modest level of horizontal differentiation. Except for the highest management level, the span of control was narrow with one or two subordinates reporting to a single manager or supervisor. However, the use of indirect reporting lines directly to the highest level of management suggested a flat organization structure with short

communication channels. A high degree of formalization was found with the organization's service offering (i.e., highly standardized outputs), yet the organization's use of written rules and procedures was limited to encourage creativity and reflective thinking.

Because the study's focus was on assessing the practicality of the OEC in an effort to validate the tool, a conscious decision was made by the investigator to focus the metaevaluation specifically on the evaluation findings and process from the client's perspective in an effort to identify strengths and weaknesses of the OEC process. Procedural issues and opportunities that were discovered by the evaluator during the field study were systematically noted and addressed by revising or clarifying the OEC where appropriate.

*Evaluator Perspectives.* At the initial meeting with the president and a managing partner, the attending managing partner was identified as the primary liaison for ongoing communications and data requests. The information collected to establish the boundaries of the evaluation, conduct the performance needs assessment, and define the criteria of merit was obtained from personal interviews with the managing partners of the firm and document review. The organization engaged an external consulting firm to assist with strategic planning for the organization several years prior to this evaluation. As a result, the company had a well-defined mission statement and an existing SWOT (strengths, weaknesses, opportunities, and threats) analysis. The SWOT analysis was helpful from an external evaluator's perspective, as it allowed for a quick understanding of the strengths and weaknesses of the organization. Rather than conduct a new SWOT analysis, efforts were made to identify what had changed since the analysis was completed. In this particular case, the analysis was still relevant and only required minor updates to reflect the current situation with respect to market developments, the financial health of the firm, and the economic environment. Based on this observation, the OEC

checkpoint referring to the SWOT analysis was modified to suggest that the evaluator request a copy of the most recent SWOT analysis prior to commencing the SWOT analysis. This revision to the checkpoint supported efforts to increase the parsimony of the OEC.

The universal criteria of merit acted as a catalyst to encourage and focus discussion leading to the definition of the measures and performance standards used in the evaluation. Without the explicit criteria of merit included in the OEC, the evaluation may have been severely delayed due to the organization's emphasis on an informal work environment and limited use of performance monitoring, among other contributing factors. Furthermore, the universal criteria of merit stimulated discussion on the contextual matters that define an effective organization. This discussion resulted in an expanded list of contextual criteria of merit used in the evaluation.

More guidance than expected was required for the importance weighting task as the senior managers rated nearly every criterion as "high" importance, and many were rated as "emergency." After discussion with the client, it was determined that none of the criteria was considered urgent *and* important, and adjustments were made accordingly to the importance ratings. It was also found to be of value that the senior managers all "publicly" agreed on the importance weightings as well as the criteria, measures, and performance standards. This addendum as well as clarification on the importance rating "emergency" was incorporated into the OEC.

Creating performance matrices, the final checkpoint in Step 3 of the OEC, was found to be one of the more time-consuming aspects of the evaluation process. While the entire evaluation was conducted in approximately three months, the development and approval of the performance matrices consumed nearly one-fourth of this time. Based on these findings and feedback received

from expert panel members, an evaluator-facilitated workshop to develop the performance matrices was included in the revised OEC.

*Client Perspectives.* This was the first organizational evaluation that the client had engaged in, excluding financial audits that were conducted by a public accounting firm. To facilitate discussion regarding the evaluative conclusions, a written evaluation report was provided to the client one week prior to the oral presentation of the results. Following the presentation of the evaluation findings, an interview with two of the managing partners was conducted to assess the evaluation's validity, credibility, utility, cost-effectiveness, and ethicality. The interview with the evaluation client was guided by six open-ended questions, including:

1. Did you find the evaluative conclusions to be credible? Why or why not? How could the credibility be improved?
2. Did you find the evaluative conclusions to be valid? Why or why not? How could the validity be improved?
3. In terms of time, effort, and disruption to the organization, was the evaluation cost-effective? Why or why not?
4. To what extent did you find the evaluation to be fair and complete, ethical, and show due regard for those involved in the evaluation?
5. Overall, to what extent did you find the evaluation to be useful? Why or why not? How could the utility be improved?
6. What actions, if any, do you expect the organization will take because of this evaluation?

Overall, the client indicated that the results confirmed what was believed to be deficient areas or strengths of the organization. With respect to the evaluation's credibility, the client found favor with the use of "hard data" such as financial ratios and the use of sound techniques to collect data. The inclusion of quantitative data was highly appealing to one of the interview

participants, while the other tended to favor thick description along the line of storytelling. When asked specifically about the validity of the findings, the client noted the completeness of the evaluation findings as illustrated by the comment, “I was surprised it picked up as much as it did. I did not think there was so much information to work with.” In terms of time, effort, and disruption to the organization, the client suggested there was “very low impact on the organization” and a “low cost for what was produced.”

With respect to the fairness of the evaluation, the client indicated that the process was transparent and that the evaluation was impartial. When asked about the utility of the evaluation, both interview participants commented on its relevance and indicated the organization would integrate several of the metrics included in the evaluation into their monitoring and performance measurement system. In addition, the recommendations were of considerable interest and the participants expressed their intent to act on them. Based on the high level of interest shown by the evaluation client in the recommendations included in the written and oral reports, several paragraphs were added to the OEC to address the limitations and consequences of including recommendations in an evaluation report.

A deficiency in the evaluation findings noted by the client was the exclusion of any reference to the organization’s inability to focus its efforts. More specifically, the organization perceived its ability to execute a plan as weak, and this absence was not identified during the evaluation. This weakness would have been most likely to have been discovered during the SWOT analysis in Step 2 of the OEC. This oversight provides an example of potential evaluation myopia that may have resulted from rigidly following a checklist and not reflecting beyond the checklist.

Based on the findings from this case study, numerous points of clarification were added to the OEC as noted in this section. Elements of the OEC that were reinforced by this case study included the need for multiple measures, transparency in the development of the performance matrices, brevity and clarity in reporting, and the importance of considering latent contextual variables (e.g., organizational culture) during the evaluation process.

#### Aspects of Reliability and Validity

Several approaches were used to minimize measurement error in this validity study. The first was the use of a pilot test of the critical feedback survey to assess the clarity of directions, question wording, and rating procedure, among other elements (Fink, 2003). The use of a systematic approach for the expert panel members to apply the ratings also contributed towards minimizing measurement error. Internal consistency of the survey instrument was estimated using Cronbach's coefficient alpha. In this particular study, coefficient alpha ranged from .79 to .96 suggesting a high estimate of reliability based on the mean inter-item correlation. What's more, the use of expert panel members with diverse backgrounds suggests potential stability over time (Yin, 2009). With respect to the content analysis, the unit of analysis was a character string that referred to a specific idea (e.g., a word). This is "the smallest and, as far as reliability is concerned, the safest recording unit for written documents" (Krippendorff, 2004, p. 104).

The aspects of construct validity addressed in this investigation included content, substantive, external, and consequential. The content aspect of construct validity includes evidence of content relevance and representativeness. Both of these elements were appraised by the expert panel members using the critical feedback survey and their professional judgments captured in the written comments made directly on the OEC document. Substantive validity, the extent to which items in the OEC can be accounted for specifically in the context of the

environment (Loevinger, 1957; Messick 1995), was addressed by the expert panel review and through the field study where the contextual element was prominent.

The external aspect of construct validity was considered in terms of the extent to which the OEC was consistent with other empirically validated models to assess organizational effectiveness. As part of the checklist development process, six of the most influential and widely accepted approaches to assess organizational effectiveness were identified (Martz, 2008). Of these six approaches, only Quinn and Rohrbaugh's (1981, 1983) competing values framework (CVF) specifically included criteria of merit for evaluating organization effectiveness and was empirically validated using multitrait-multimethod analysis and multidimensional scaling (Quinn & Spreitzer, 1991) and structural equation modeling (Kalliath, Bluedorn, & Gillespie, 1999). The universal criteria of merit embedded in the OEC are consistent with the criteria used in the CVF to assess organizational effectiveness providing support for external validity of the OEC.

The final facet of validity addressed in this investigation was consequential validity. That is, providing evidence and rationale for evaluating the intended and unintended consequences of interpretation and use (Messick, 1995). Because the OEC is intended to benefit users (e.g., professional evaluators, organizational consultants, managers) and the organizations that are evaluated, it is important to accrue evidence of such positive consequences. Overall, the expert panel members provided positive feedback regarding the potential utility of the OEC in the critical feedback survey. Table 3 summarizes the expert panel ratings for each criterion used for evaluating the checklist. The highest rating was found with the fairness criterion ( $M = 7.05$ ,  $SD = 1.52$ ) followed by concreteness ( $M = 6.96$ ,  $SD = 1.69$ ), clarity ( $M = 6.93$ ,  $SD = 1.59$ ), and pertinence to the content area ( $M = 6.86$ ,  $SD = 1.69$ ). The field study provided evidence of the OEC's potential utility from the evaluator's and organization's perspectives by demonstrating its

practicality and feasibility in application. In addition, the organization's assessment of the evaluation process was positive as indicated by the favorable responses collected during the post-evaluation interview. Together, the expert panel feedback and the case study evaluation findings provided evidence and rationale for consequential validity of the checklist.

Table 3: Critical feedback survey ratings

	<i>M</i>	<i>SD</i>	$\alpha$
Applicability	6.52	1.79	.83
Clarity	6.93	1.59	.79
Comprehensiveness	6.46	1.80	.90
Concreteness	6.96	1.69	.93
Ease of use	5.61	1.68	.96
Fairness	7.05	1.52	.87
Parsimony	5.70	1.71	.80
Pertinence	6.86	1.69	.86

### Limitations

As with all evaluation checklists, a number of years of field-testing and use are required to thoroughly validate and refine the tool. Repeated use of a checklist is the preferred method for validation, and suggests a major limitation of this validation approach. Nevertheless, the multimethod design was an attempt to establish a reasonable level of validity of the new tool. It was felt that some exposure to a real-world setting was better than no exposure at all, and that feedback from targeted users and the evaluation client would provide insight for improving the

OEC. As a result, issues that may not have been foreseen in a laboratory were more likely to be discovered.

Another limitation was the evaluation client selected for the field study. Although specific selection criteria were used to qualify the organization for inclusion in the study, additional aspects of the OEC process may have been illuminated had the organization been multidivisional, larger, at a different developmental stage, or in a hypercompetitive marketplace, among a variety of other contextual differentiators. In addition, multiple field studies of organizations of varying types, structures, and purposes would have provided exposure to a broader range of organizational characteristics upon which to subject the checklist and OEC process. Consequently, the limitations attributed to using a single-case field study temper the strength of the evidence supporting consequential validity.

On a similar note, the selection of expert panel members was another limitation of this methodology. The selection was one of convenience, yet purposeful in its attempt to include a diverse array of targeted users and subject matter experts in an effort to provide evidence that the OEC can be applied to a broad variety of organizational contexts. However, only two of the expert panel members were specifically engaged in working with nonprofit organizations, and this may have limited the insights regarding the ability to generalize the OEC across organization types (e.g., profit and nonprofit). In addition, the inclusion of participants from emerging economies may have offered distinct perspectives pertinent to using the OEC to evaluate organizations operating in developing regions that were not identified by the participants in this study.

## Conclusions

The validation of a checklist plays an important role in establishing the credibility and utility of the checklist—particularly when the checklist is used for evaluative purposes. This paper outlined a two-part validation process of an evaluative checklist that featured a mixed methods approach using survey research and case study research. The validation method proposed in this article is structured, yet flexible, and can be applied to a variety of other evaluative checklists. The use of subject matter experts and targeted users combined with an actual field study provided valuable perspectives for developing, refining, and validating the OEC. As noted earlier, critical feedback based on repeated use of the checklist offers a preferred method for validating and improving an evaluative checklist. Nonetheless, the validation approach presented here is relatively quick and was demonstrated to be feasible on a limited budget all the while providing a reasonable level of validation for the OEC.

## Appendix A

### Outline of the OEC

1. Establish the boundaries of the evaluation.
  - 1.1 Identify the evaluation client, primary liaison, and power brokers.
  - 1.2 Clarify the organizational domain to be evaluated.
  - 1.3 Clarify why the evaluation is being requested.
  - 1.4 Clarify the timeframe to be employed.
  - 1.5 Clarify the resources available for the evaluation.
  - 1.6 Identify the primary beneficiaries and organizational participants.
  - 1.7 Conduct an evaluability assessment.
  
2. Conduct a performance needs assessment.
  - 2.1 Clarify the purpose of the organization.
  - 2.2 Assess internal knowledge needs.
  - 2.3 Scan the external environment.
  - 2.4 Conduct a strength, weakness, opportunity, and threat analysis.
  - 2.5 Identify the performance-level needs of the organization.
  
3. Define the criteria to be used for the evaluation.
  - 3.1 Review the universal criteria of merit for organizational effectiveness.
  - 3.2 Add contextual criteria identified in the performance needs assessment.
  - 3.3 Determine the importance ratings for each criterion.
  - 3.4 Identify performance measures for each criterion.
  - 3.5 Identify performance standards for each criterion.
  - 3.6 Create performance matrices for each criterion.
  
4. Plan and implement the evaluation.
  - 4.1 Identify data sources.
  - 4.2 Identify data collection methods.
  - 4.3 Collect and analyze data.
  
5. Synthesize performance data with values.
  - 5.1 Create a performance profile for each criterion.
  - 5.2 Create a profile of organizational effectiveness.
  - 5.3 Identify organizational strengths and weaknesses.
  
6. Communicate and report evaluation activities.
  - 6.1 Distribute regular communications about the evaluation progress.
  - 6.2 Deliver a draft written report to client for review and comment.
  - 6.3 Edit report to include points of clarification or reaction statements.
  - 6.4 Present written and oral reports to client.
  - 6.5 Provide follow-on support as requested by client.

## References

- Altschuld, J. W., & Zheng, H. Y. (1995). Assessing the effectiveness of research organizations: An examination of multiple approaches. *Evaluation Review, 19*, 197-216.
- Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*, 205-215.
- Cooksy, L. J., & Caracelli, V. J. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Evaluation, 26*, 31-42.
- Corbin, J. M., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles: Sage.
- Coryn, C. L. S. (2007). *Evaluation of researchers and their research: Toward making the implicit explicit*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.
- Fink, A. (2003). *The survey handbook* (2nd ed.). Thousand Oaks, CA: Sage.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*, 255-274.
- Hales, B., Terblanche, M., Fowler, R., & Sibbald, W. (2008). Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care, 20*, 22-30.
- Hales, B. M. & Pronovost, P. J. (2006). The checklist: A tool for error management and performance improvement. *Journal of Critical Care, 21*, 231-235.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*, 1277-1288.
- Kalliath, T. J., Bluedorn, A. C., & Gillespie, D. F. (1999). A confirmatory factor analysis of the competing values instrument. *Educational and Psychological Measurement, 59*, 143-158.
- Keil, M., Li, L., Matthiassen, L., & Zheng, G. (2008). The influence of checklists and roles on software practitioner risk perception and decision-making. *Journal of Systems and Software, 81*, 908-919.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory [Monograph Supplement 9]. *Psychology Reports, 3*, 635-694.

- Martz, W. (2008). Evaluating organizational effectiveness. Dissertation Abstracts International, 69 (07). Publication No. ATT3323530.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50, 741-749.
- Owen, J. M., & Lambert, F. C. (1998). Evaluation and the information needs of organizational leaders. *American Journal of Evaluation*, 19, 355-365.
- Papadimitriou, D. (2007). Conceptualizing effectiveness in a non-profit organizational environment: An exploratory study. *International Journal of Public Sector Management*, 20, 571-587.
- Persaud, N. (2007). *Conceptual and Practical Analysis of Costs and Benefits: Developing a Cost-Analysis Tool for Practical Program Evaluation*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.
- Quinn, R. E., & Rohrbaugh, J. (1981). A competing values approach to organizational effectiveness. *Public Productivity Review*, 2, 122-140.
- Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science*, 29, 363-377.
- Quinn, R. E., & Spreitzer, G. M. (1991). The psychometrics of the competing values culture instrument and an analysis of the impact of organizational culture on quality of life. *Research in Organizational Change and Development*, 5, 114-142.
- Ridley, C. R., & Mendoza, D. W. (1993). Putting organizational effectiveness into practice: The preeminent consultation task. *Journal of Counseling and Development*, 72, 168-178.
- Sanders, J. R. (2001). A vision for evaluation. *American Journal of Evaluation*, 22, 363-366.
- Schroeter, D. C. (2008). *Sustainability evaluation: Development and validation of an evaluation checklist*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.
- Schweigert, F. J. (2006). The meaning of effectiveness in assessing community initiatives. *American Journal of Evaluation*, 27, 416-436.
- Scriven, M. (1994). The final synthesis. *Evaluation Practice*, 15, 367-382.
- Scriven, M. (2005). Checklists. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 53-59). Thousand Oaks, CA: Sage.
- Scriven, M. (2007). *The logic and methodology of checklists*. Retrieved December 30, 2007 from [http://www.wmich.edu/evalctr/checklists/papers/logic&methodology\\_dec07.pdf](http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_dec07.pdf)

- Stufflebeam, D. L. (2000, July) *Guidelines for developing evaluation checklists: The checklist development checklist*. Retrieved April 22, 2007 from [http://www.wmich.edu/evalctr/checklists/guidelines\\_cdc.pdf](http://www.wmich.edu/evalctr/checklists/guidelines_cdc.pdf)
- Stufflebeam, D. L. (2001). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22, 71-79.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.