

GUIDELINES FOR DEVELOPING EVALUATION CHECKLISTS: THE CHECKLISTS DEVELOPMENT CHECKLIST (CDC)

Daniel L. Stufflebeam
July 2000

This paper is intended to provide practical guidance to persons desiring to develop a checklist as a tool for evaluating in a particular area. Checklists are valuable evaluation devices when carefully developed, validated, and applied. A sound evaluation checklist clarifies the criteria that at least should be considered when evaluating something in a particular area; aids the evaluator not to forget important criteria; and enhances the assessment's objectivity, credibility, and reproducibility. Moreover, such a checklist is useful in planning an enterprise, monitoring and guiding its operation, and assessing its outcomes. In the evaluation vernacular, checklists are useful for both formative and summative evaluations.

The guidelines that follow in this paper are based on the author's 30+ years' experience in developing and applying evaluation checklists. These include checklists to guide the planning and implementation of program and personnel evaluations and metaevaluations, an evaluation contracting checklist, and professional standards for both program and personnel evaluations. All of these checklists and standards (Joint Committee, 1981, 1988, 1994) have been developed, applied, refined, published, and used fairly widely. Copies of some of these checklists and others—especially those constructed by my colleague Michael Scriven—are presented on the Checklist Project web site (www.wmich.edu/evalctr/checklists/).

The remainder of this paper includes a synthesized representation of processes used to develop the above checklists and standards. Essentially, what follows is a checklist for developing checklists. Basically, the presented guidelines are divided into 12 main checkpoints, with each further divided into several more specific checkpoints. At this point in the development of the CHECKLISTS DEVELOPMENT CHECKLIST (CDC), no specific scoring procedure is provided or deemed necessary. It is hoped that users will find the CDC helpful in planning and proceeding through the steps required to develop and employ a sound checklist. Definitions and rationales for each main checkpoint appear in the Appendix. Without further introduction, the remainder of this paper presents the CDC.



CHECKLISTS DEVELOPMENT CHECKLIST (CDC)

1. Focus the checklist task

- Define the content area of interest
- Define the checklist's intended uses
- Reflect on and draw upon pertinent training and experience
- Study the relevant literature
- Engage and have conversations with experts in the content area
- Clarify and justify the criteria to be met by the checklist (e.g., pertinence, comprehensiveness, clarity, concreteness, ease of use, parsimony, applicability to the full range of intended uses, and fairness)

2. Make a candidate list of checkpoints

- List descriptors for well-established criteria of merit
- Briefly define each of the initial checkpoints
- Add descriptors for checkpoints needed to round out a definition of merit for the content area
- Provide definitions for each of the added descriptors

3. Classify and sort the checkpoints

- Write each descriptor and definition on a separate 4" x 6" card
- Sort the cards in search of categories
- Identify the main candidate categories and label each category

4. Define and flesh out the categories

- Define each category and its key concepts and terms
- Write a rationale for each category
- Present relevant warnings about being overzealous in applying the checkpoint
- Review the checkpoints in each category for inclusiveness, clarity, and parsimony
- Add, subtract, and rewrite checkpoints as appropriate

5. Determine the order of categories

- Decide if order is an important consideration regarding the intended uses of the checklist
- If so, write a rationale for the preferred order
- Provide an ordering of the categories

6. Obtain initial reviews of the checklist

- Prepare a review version of the checklist
- Engage potential users to review and critique the checklist
- Interview the critics to gain an in-depth understanding of their concerns and suggestions
- List the issues in need of attention

7. Revise the checklist content

- Examine and decide how to address the identified issues
- Rewrite the checklist content

8. Delineate and format the checklist to serve the intended uses
<input type="checkbox"/> Determine with potential users whether category and/or total scores are needed or desired <input type="checkbox"/> Determine with users what needs exist regarding differential weighting of categories and/or individual checkpoints <input type="checkbox"/> Determine with users any checkpoints or categories of checkpoints that must be passed for a satisfactory score on the overall checklist <input type="checkbox"/> Determine with users what needs exist regarding profiling of checklist results <input type="checkbox"/> Format the checklist based on the above determinations
9. Evaluate the checklist
<input type="checkbox"/> Obtain reviews of the checklist from intended users and relevant experts <input type="checkbox"/> Engage intended users to field-test the checklist <input type="checkbox"/> Generally, assess whether the checklist meets the requirements of pertinence, comprehensiveness, clarity, applicability to the full range of intended uses, concreteness, parsimony, ease of use, and fairness
10. Finalize the checklist
<input type="checkbox"/> Systematically consider and address the review and field-test findings <input type="checkbox"/> Print the finalized checklist
11. Apply and disseminate the checklist
<input type="checkbox"/> Apply the checklist to its intended use <input type="checkbox"/> Make the checklist available via such means as journals, professional papers, web pages, etc. <input type="checkbox"/> Invite users to provide feedback to the developer
12. Periodically review and revise the checklist
<input type="checkbox"/> Use all available feedback to review and improve the checklist at appropriate intervals

Closing

This paper has presented the first, very preliminary edition of the CDC. It is hoped that the presented draft guidelines, even in their primitive form, will be useful to persons needing to develop and apply evaluation checklists. Of course, the CDC itself needs fuller processing through all the CDC steps. Users and reviewers of this first draft can facilitate the CDC's further development and validation of the CDC by sending in their criticisms and suggestions¹. All such input will be appreciated and considered in the further development of the CDC. The ultimate test of the CDC will be its impact on helping evaluators to develop and employ outstanding checklists.

¹ Send feedback via email to Daniel.Stufflebeam@wmich.edu; or mail to The Evaluation Center, Western Michigan University, Kalamazoo, MI 49008; or fax to 616-387-5923.

References

- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards. How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage.

Appendix: Definitions and Justifications for Checkpoints

1. Focus the checklist task

The first main checkpoint—Focus the checklist task—is a key foundational step.

Explanation. This checkpoint requires clear definition of the types of programs, services, personnel, or other objects to be evaluated and of the checklist's intended users. It requires the checklist developer to identify and then bring to bear an appropriate knowledge base, including personal training and experience, review of relevant literature, and involvement of pertinent experts.

The checkpoint also requires clarification and justification of the criteria to be met by the checklist. At a minimum, these criteria include pertinence of the specific checkpoints to the content area; comprehensiveness in including all the important checkpoints; clarity; applicability to the full range of intended uses; sufficient concreteness for application; parsimony aimed at minimizing repetition or overlap of checkpoints; ease of use; and fairness, especially to avoid evoking unreasonable expectations. Just as the author has determined the above criteria based on much experience, the developers of other checklists will often want to add additional criteria for evaluating the checklist based on relevant experience, input from others, and study of the particular content area. Justification for the checklist soundness criteria will often be self-evident and readily accepted by users, as I believe to be the case with those listed above. Nevertheless, it is often wise to put the criteria up for review, critique, and suggestions and to improve them accordingly.

Rationale. Checklist developers must establish a sound foundation for the intended checklist. Only then can the checklist be specifically targeted, coherent, possessing of integrity, valid, credible, and helpful to an identified constituency.

2. Make a candidate list of checkpoints

This step is a quite creative and exploratory activity. It is aimed at drafting a starter list of checkpoints.

Explanation. One can begin by listing descriptors of well-established criteria in the area of interest. Familiar examples in the area of educational and psychological measurement are measurement validity and reliability. This initial identification step can be followed by providing brief definitions of each descriptor, e.g., do the following—so that the following desired outcome —will be achieved. Following this opening activity one should try to add all the descriptors and definitions needed to round out a definition of merit for the content area. In this process one need not worry about order of the growing list of checkpoints.

Rationale. This early step is important to get the process of producing the needed content for the checklist started. It helps both to incorporate well-known, pertinent criteria and to create a more extensive list. At this early stage, hard thinking and unencumbered creativity are required. There is no need here to burden the process with worries about categorizing checkpoints, weighting, developing a scoring scheme, etc. What is important is to generate a useful working list of checkpoints with descriptors and associated definitions.

3. Classify and sort the checkpoints

Given an initial list of checkpoints, a next useful activity can involve classifying and sorting the checkpoints.

Explanation. After writing an initial list of more or less randomly ordered checkpoints on a tablet or computer, I have often found it useful to enter each checkpoint (including the descriptor and definition) on a separate 4" x 6" index card. This facilitates sorting the checkpoints in search of categories and provides space for recording pertinent notes that will be useful later about such matters as rationale, relevant cases, order, caveats, etc. Ultimately, this sorting task culminates in the identification of main candidate categories.

Rationale. While long lists of checkpoints can be useful for such tasks as packing a suitcase, typically even such a simple checklist becomes more useful when like items are grouped together. This is especially so when the items can be grouped by diagnostic area or function. Such groupings of checkpoints help evaluators not just to show particular items of strength and weakness, but larger areas that should be improved or reinforced. Grouping the checkpoints also helps the checklist developer to see gaps and areas of overlap, which is important for expanding and refining the checklist. Moreover, categorized checkpoints facilitate the scoring process, should one be needed.

4. Define and flesh out the categories

Once the checkpoints have been categorized, it is important to carefully define and flesh out the categories.

Explanation. Each category and its key concepts should be defined. The importance of the category should also be justified. To assure that the category is treated in a balanced way, it can also be important to include caveats. These especially include warnings against concentrating too much on a given category of checkpoints without simultaneously considering the rest of the checklist. Following the write-up of each category, the checklist developer should review and assess the category's checkpoints. This is a good time to add, subtract, and rewrite the checkpoints in order to assure that each category is fully defined, clear, and efficient in its coverage.

Rationale. Unless one seriously defines and improves the categories of checkpoints, they are unlikely to withstand scrutiny or be maximally useful. This appendix is an attempt to model what is involved in this process.

5. Determine the order of categories and of checkpoints within categories

Once the checkpoints have been grouped, a determination should be made regarding the ordering of the categories.

Explanation. The first step in applying this checkpoint involves deciding to what extent order is an important consideration. Obviously, some order will be decided for all checkpoints and categories of checkpoints, but a particular sequence of checklist categories and individual checkpoints can be important in some areas, such as medical diagnosis, planning evaluation studies, and checking an airplane's readiness to fly. That said, it should also be noted that even in specifically ordered checklists, application of the checklist items will often be looping, repetitive, and iterative.

In making the determination of sequence, one should review the checklist's intended uses. Key considerations include whether certain checkpoints depend on a previous application of other checkpoints in given uses of the checklist or whether some checkpoints would point up the appropriateness of proceeding with or aborting an activity before even considering other checkpoints. The first of these points refers to the checklist's effectiveness while the latter pertains to its contributions to efficiency. An example of the latter is seen in *The Program Evaluation Standards* (Joint Committee, 1994) where the Utility category precedes the Feasibility, Propriety, and Accuracy categories. If, at the outset of planning an evaluation, one can see that its findings would not be used, then the evaluation would best be terminated. In such a case, the Utility standards are of first order concern and there would be no need to work through the Feasibility, Propriety, and Accuracy categories.

Once the categories have been ordered, the checklist developer should assure that the checkpoints in each category are sequenced logically and functionally. Essentially, this step involves hard thinking and testing the logic and coherence of different sequences of checkpoints. Getting the views of a few intended users can be helpful in this step.

Rationale. As the above discussion shows, ordering checklist categories is an important consideration. A logical, functional order helps users to proceed efficiently through activities on which the checklist is focused and successively to build on the results obtained with prior checkpoints.

6. Obtain initial reviews of the checklist

Once the checklist categories and individual checkpoints have been appropriately sequenced, the checklist developer should obtain an initial set of reviews.

Explanation. At the outset, the checklist developer should prepare a review version of the checklist. Potential users should then be recruited and engaged to provide written, critical reviews of the checklist. They should be asked to mark up the checklist itself and to supplement this with written critical comments and suggestions. After carefully considering the critical feedback, the checklist developer should interview those respondents who provided especially useful and/or unclear feedback. Such interviews can add important information and increase the checklist developer's insights into what was provided. In rounding out the initial review stage, the checklist developer is advised to list all the issues that require attention. This list provides an agenda for improving the checklist.

7. Revise the checklist content

In the overall scheme of checklist development, improvement of the checklist is an ongoing process. The first occasion of this occurs following the initial independent review of the checklist.

Explanation. The checklist developer should carefully reflect on each issue surfaced in the independent review and decide whether and, if so, how the checklist should be changed to ameliorate each identified problem. Making such changes must follow a thoughtful process, taking into account how the different, potentially conflicting changes can be brought into harmony. The checklist developer should make sure that the changes in combination are mutually reinforcing and make sense for improving the checklist as a whole. Also, the checklist developer should sustain and build on the checklist's strengths. Following this general approach, the checklist developer should rewrite the checklist as appropriate.

Rationale. The acquisition of independent feedback is important for improving evaluation checklists, but such feedback is worthwhile only if it is seriously considered and appropriately applied. I have found it useful to make lists of all obtained criticisms, suggestions, and areas of strength; make notes on whether and how to make improvements regarding each one; and then to rewrite the checklist, taking into account the whole set of information.

8. Delineate and format the checklist to serve the intended uses

Following the preceding systematic processes to arrive at a comprehensive and usefully organized and sequenced set of checkpoints, the checklist developer can take a further step to make the checklist useful. This step involves formatting the checklist to serve particular users and uses.

Explanation. At this stage of developing the checklist, the checklist developer and intended users can usefully consider a range of options that might enhance the checklist's utility. One possibility involves steps to compute and assign value judgments to scores for checklist categories and/or the overall checklist. Examples of how to do these may be seen in the two checklists on the Checklist Project web site (www.wmich.edu/evalctr/checklists/) that address individual personnel evaluations and personnel evaluation systems.

An associated option is to determine differential weights for individual checkpoints or categories of checkpoints. In my experience, the gain from weighting individual checkpoints is not worth the investment of time, effort, and cost required to do it right. A better approach is to weight categories of checkpoints either by varying the number of checkpoints assigned to categories according to the judged relative importance of the categories or to standardize category scores and multiply them by their respective weights. A score across all the categories can be obtained by totaling the unweighted or weighted category scores or doing both and dividing the number of categories. These then can be interpreted by comparing them with ranges of scores predetermined to correspond to such judgments as very poor, poor, good, very good, and excellent. One way of doing this can be seen by referring to the checklists on the Checklist Project web site (www.wmich.edu/evalctr/checklists/) that are concerned with individual personnel evaluations and personnel evaluation systems.

However one proceeds with the issue of weighting, it is important to consider that satisfaction or failure of checkpoints or categories of checkpoints judged to be essential should override any weight and sum score. For example, in professional measurement, it matters not how well a test scores on all the other criteria if it fails the crucial requirement of validity.

Beyond dealing with the scoring and weighting issues, it is sometimes useful to provide steps for profiling the results of applying the checklist. Bar graphs can be especially useful for this purpose. Examples of bar graphs used to profile findings from applying checklists can be seen in any issue of *Consumer Reports* magazine.

The final step in this stage of checklist development is to put the checklist in a user friendly format. The foundation for this is set by systematically attending to the foregoing formatting items. At this stage, it can also be useful to obtain assistance from a graphics expert.

Rationale. Formatting a checklist for sound and relatively easy application is essential. Without this, the checklist likely will have only academic merit. Clearly, it is important when refining the checklist to keep front and center the intended users and intended uses. The checklist should be set up to facilitate reaching valid evaluative conclusions. It should meet the requirements of the users and yet be kept appropriately simple in form and procedures.

9. Evaluate the checklist

Having executed the preceding eight steps—which collectively are designed to produce an appropriately targeted and fully functional checklist—the checklist developer is next ready to thoroughly evaluate the checklist.

Explanation. This comprehensive evaluation should include at least three parts. First, the checklist developers should engage intended users and relevant experts to review and provide their written critiques of the checklist. The critiques should include both marked-up copies of the checklist and supplementary written comments. The second part of the evaluation should engage intended users to apply the checklist to the range of intended uses, e.g., both formative and summative evaluations. The third step involves summing up the merit and utility of the checklist based on the reviews and field tests and diagnosing deficiencies that need to be addressed.

All three evaluative steps should be keyed to a common set of criteria. At a minimum, it is recommended that these criteria include pertinence to the content area, comprehensiveness, clarity, applicability to the full range of intended uses, concreteness, parsimony, ease of use, and fairness.

Rationale. Before applying a checklist to its primary intended use and especially before disseminating it for widespread application, it is important to subject it to as much review and field-testing as is practicable. This can be a fairly concentrated formal evaluation or an ongoing iterative process. The evaluations provide both assurances regarding the checklist's quality and direction for improvement.

10. Finalize the checklist

Kurt Lewin once wrote that the improvement process involves unfreeze, move, refreeze, unfreeze, etc. That is the sense in which checklist developers should consider the tenth step of finalizing the evaluation checklist. Practically, one needs to draw together the body of relevant evaluative feedback at some point and prepare the checklist for application. But, as is the case in manufacturing automobiles, further feedback from actual use and advancements in the relevant content should be used down the road to open up the checklist to further assessment, development, and improvement.

Explanation. There are two main steps in finalizing a checklist. The first involves systematically considering and addressing the review and field-test results. In rare cases, the feedback may be so negative that the checklist development effort should be aborted or completely recycled. If so, so be it. However, if the preceding nine recommended steps have been carefully followed, the main checklist development task usually will involve refinements.

After the checklist has been judged basically acceptable, it should be finalized and printed in a form for dissemination. The printed material may include back-up material such as appears in this appendix as well as the checklist.

11. Apply and disseminate the checklist

With checkpoint 11, the checklist developer reaches the stage of application and sharing.

Explanation. The checklist will first be applied to its primary intended use. Simultaneously or later, the checklist developer may decide to make the checklist available for wider use. Dissemination can be accomplished through a variety of means—journal articles, professional papers, web pages, conference presentations, workshops, etc. Whenever one disseminates a checklist, it is wise to invite feedback describing and assessing the applications.

Rationale. Clearly, the point of checklist development is to serve the predetermined use. Also, the checklist developer can provide a valuable service to professional colleagues by making a sound checklist available for their use. It is always desirable to invite users to provide critical feedback, since checklist development is an ongoing process.

12. Periodically review and revise the checklist

Every application of the checklist and every invited review of a checklist provides information that may be useful in improving the checklist.

Explanation. The checklist developer should value and maintain a file of such information. At appropriate intervals, the checklist developer should review and address the issues found in this information. The key steps are to invite case descriptions and critiques, systematically file the information, and periodically review and use the information for improving the checklist. While no set timetable for this work is recommended, it can be beneficial to perform the review and revision work about annually.

Rationale. As the technology of evaluation and the targeted content area for a checklist develops, a checklist sooner or later is likely to be out of date. Also, applications of the checklist will often point up areas for improvement. Thus, once again, it is emphasized that checklist evaluation and development should be an ongoing process.

This checklist is being provided as a free service to the user. The provider of the checklist has not modified or adapted the checklist to fit the specific needs of the user and the user is executing his or her own discretion and judgment in using the checklist. The provider of the checklist makes no representations or warranties that this checklist is fit for the particular purpose contemplated by user and specifically disclaims any such warranties or representations.