Development of Policy Checklists for
Student Evaluation Practices:
Application of a Checklist Development Process

(A work in progress)

Arlen R. Gullickson

The Evaluation Center
Western Michigan University

There is clear, strong consensus that teachers' evaluations of students should be improved. Study after study has identified weaknesses in teachers' knowledge, skills, and practices pertaining to student evaluations. These studies date back more than 30 years, yet the problem persists. Part of the problem stems from the fact that teachers and administrators have minimal training in how to conduct student evaluations (Gullickson, 1993; Mayo, 1967; Sanders & Vogel, 1993; Shafer & Lissitz, 1987; Stiggins, 1991; Stiggins & Conklin, 1992; Wolmut, 1988). Part results from the substantial complexity of the process (Crooks, Kane, & Cohen, 1996; Shepard, 1989, Brandt, 1989; Gallagher, 1995; Teeter, 1995; Wagner, 1995; Kelley, 1989; Hills, 1991; and Wiggins, 1994).

Checklists appear to be a means to help educators address the complexity of student evaluations and improve their skills. Well-formed checklists can identify the range of conditions and procedures that should occur in well-constructed and conducted evaluations. Checklists also can guide teachers and others as they review and work to improve their evaluation practices.

This paper presents two draft checklists that can be used to improve student evaluation practices. I present the checklists not as finished products, but as products in the making. The checklists are intended to help teachers and schools improve conditions for and ultimately the student evaluations themselves. As such, the paper responds to substantial concerns about the quality and calls for improvement of student evaluations in our nation's classrooms. Yet, as you will see, the paper focuses as much on the methodology used to prepare the checklists. I do this in the hope that others (e.g., teachers and administrators) will apply these same methods to prepare their own checklists to serve local student evaluation needs.

The checklists presented here stem from two opportunities that meld together in significant and useful ways. As most of you know, I work at a place called The Evaluation Center at Western Michigan University. The Center's mission is "to provide national and international leadership for advancing theory and practice of program, personnel, and student/constituent evaluation, as applied primarily to education and human services." That mission makes it possible for me to engage in an array of interesting evaluation activities that serve educators.

The first opportunity occurred in 1996 when we received funding from the National Science Foundation for a project to enhance evaluation capacity in the mathematics, engineering, and technology education communities through materials development, training, and support services. That grant gave the Center support to explore and conduct a variety of work. For example, we host a summer institute on evaluation. One of the best aspects of the grant is that it affords me the opportunity to engage some of the country's top evaluators to think about, explore, and share some of their ideas. In 2000 the grant brought Drs. Stufflebeam and Scriven together to think about and develop checklists to serve evaluators.

That checklist project continues today, but its initial efforts resulted in two especially important documents: *The Logic and Methodology of Checklists* (Scriven, 2000) and the *Checklists Development Checklist* (Stufflebeam, 2000). Those two documents succinctly describe the rationale, importance, and utility of checklists, and provide clear-cut directions on how to develop a strong checklist. The documents are brief, easy to read, and can be readily employed by persons who are not experts in evaluation. These characteristics are particularly important for teachers, administrators, and students, who may not be skilled evaluators but desire to construct their own checklist tools for evaluation purposes.

The second important opportunity occurred because The Evaluation Center serves as the home base for the Joint Committee on Standards for Educational Evaluation (Joint Committee). As stated on the Joint Committee's home page (http://jc.wmich.edu), the Joint Committee, created in 1975, is a coalition of major professional associations concerned with the quality of evaluation. The Joint Committee has published two sets of standards for evaluations that are now widely recognized, *The Personnel Evaluation Standards* (1988), and *The Program Evaluation Standards* (2nd edition) (1994).[1] Currently, I serve as the Committee's chair, and it is developing a third set of standards to guide teachers' evaluations of students. It is that third set of standards that prompted this evaluation checklist work.

In the remainder of this paper I briefly present my development of a checklist for student evaluations. It is a work in progress. As will become clear, the checklist focuses on what I consider to be the beginning point for student evaluations in the classroom. Important as this beginning point is, it addresses only a small portion of the work important to conducting sound student evaluations. The draft version of *The Student Evaluation Standards* provided the basic content for the checklist. In turn *The Logic and Methodology of Checklists* (Scriven, 2000) and the *Checklists Development Checklist* (Stufflebeam, 2000) provide both a foundation for understanding and practical steps for constructing the checklist.

Scriven (2000) defines *checklist* in the following way: "A checklist is taken here to be a list of factors, properties, aspects, components, criteria, tasks, or dimensions, the presence or amount of which is to be separately considered, in order to perform a certain task." He goes on to identify and describe various types of checklists: laundry list, sequential checklist, weakly sequential checklist, iterative checklist, diagnostic checklist, and the criteria of merit checklist (comlist). Scriven notes that the comlist is the checklist of choice for evaluating evaluations–the task undertaken here. In his description of comlists Scriven (2000) states,

---

[1] The Joint Committee is accredited by the American National Standards Institute (ANSI). Standards approved by ANSI become American National Standards

> The comlist is often a tough item to develop and validate: it has to meet some stringent requirements that do not apply to the simpler types of checklist discussed so far. For example, it is essential that it be complete, or very close to it, i.e., that it include every significant criterion of merit. Otherwise, something that scores well on it may be quite inferior because of its poor performance on some missing but crucial dimension of merit. Again, the criteria in a comlist should not overlap if it is to be used as a basis for scoring, to avoid "double counting" of the overlap area.

Based on the criteria Scriven provided and the fact that I wanted my checklist to serve evaluations of student evaluation practices, I chose to develop a comlist. I then turned to the *Checklists Development Checklist* (CDC) (Stufflebeam, 2000) as a tool to create the checklist. The CDC sets forward 12 primary checkpoints or practical steps for the development of a checklist.

This paper follows that CDC outline to described my efforts to develop a checklist and lists his major checkpoints in italics. This strategy is employed partly to help readers follow and understand the work completed. More importantly, the strategy shows the relevance of the CDC for groups such as teachers or administrators who wish to develop their own checklist to address student evaluation matters of importance to them.

☒　　　**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)*

The content area of interest is teachers' evaluations of students in classroom contexts. The initial intention was use by teachers in the classroom to improve their evaluation practices. That intention grew out my personal research and my extensive work with the Joint Committee in its efforts to develop clear, strong standards to guide teachers' evaluations of students to ensure that their evaluation practices uphold four basic principles of evaluation: propriety, utility, feasibility, and accuracy. Through extended discussions with and written input from experts and practitioners over a period of several years, substantial ground work had been completed to clarify the issues and justify the key points of interest. These points

were available to the checklist work as 29 standards and their associated guidelines and common errors in the draft book *The Student Evaluation Standards*.

The Joint Committee's work and the many documents it draws from suggest that evaluations of students in the classroom serve two primary purposes. First and foremost, student evaluations must provide formative assistance, guiding the learning of students by helping students and teachers see strengths, weaknesses, and avenues for growth. Secondly, student evaluations must provide accountability (summative) information regarding student progress to students, parents, and others.

The Student Evaluation Standards focus directly on criteria for evaluating student evaluations and divide them according to the four previously mentioned principles (propriety, utility, feasibility, and accuracy). The standards themselves do not require that evaluations follow any prescribed process. However, these standards do reference four general procedures essential for strong student evaluations in the classroom. These four are listed below in stepwise order.

1. Development and dissemination of school policies to guide student evaluation practices in the classroom (Standard P2 Appropriate Policies and Procedures)
2. Preparation and distribution of information by the classroom teachers to explain the teacher's expectations of students, student evaluation policy as it will be employed in the classroom, and the teacher's procedures for conducting student evaluations (Standards A1 Defined Expectations for Students and A3 Described Procedures).
3. Preparation and use of assessments to make formative and summative evaluative determinations.[2]
4. Evaluation reporting and follow-up to ensure that the evaluation effectively guides learning and accountability for students. (See, for example, U6 Effective Reporting, U7 Evaluation Follow-Up, and U12 Metaevaluation)

In addition to being ordinal in nature, these steps are expected to be coherent, each building upon the previous step. These four steps provided the beginning structure for an evaluation checklist. Important, though not immediately obvious is the fact that this structure flows out of the substantial interaction among Joint Committee members and feedback provided to the Joint Committee by reviewers and field trials of the draft standards.

☒ **2. *Make a candidate list of checkpoints***

---

[2]The majority of standards presented in *The Student Evaluation Standards* address this step. See the list of standards appended to this paper.

- *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*

Each standard presented in *The Student Evaluation Standards* includes a list of guidelines and common errors. Taken together, the guidelines and common errors provide a comprehensive and somewhat overlapping set of well-established descriptors for conducting student evaluations. As a preliminary step, I converted those descriptors to checkpoints. This resulted in a LARGE list of 400 checkpoints.

This large list made two things apparent. First, the list graphically presents the extensive and complex nature of student evaluations. Second, it suggests that student evaluation practices *in toto* are too large to be covered in a single evaluative checklist. To reduce the list in a viable way I returned to the first checkpoint in the CDC and narrowed the content area to that of policies that drive student evaluations. This provides a focus on the first two steps of the student evaluation process.

☒ **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*

Rather than employ 4" x 6" cards, I placed the descriptors in a spreadsheet and used spreadsheet fields to classify each item regarding whether it addressed policy and whether the focus was schoolwide or teacher/classroom based. Also, because several standards included items pertinent to policies, I kept a field to specify which standard was served (e.g., Standard P2 Appropriate Policies and Procedures). This field designator is kept in the checklist to encourage users of the checklist to also read associated material in *The Student Evaluation Standards.*

☒ **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*

As noted above, I made a preliminary decision to focus on student evaluation policies. The rationale for this focus stems from five important aspects. First, there is substantial literature that shows that student evaluations are not well conducted in schools. Second, school policies are viewed as the first step for conducting sound student evaluations since policies drive and guide implementation of evaluation procedures. Third, teacher policies should flow directly from and be consistent with school-based policies. Fourth, a teacher's policies should direct the teacher's procedures for conducting student evaluations. Fifth, available studies and interactions with educators indicate that schools do not have well constructed or complete policies to guide student evaluations.

The importance of school policies is set forward in the rationale for standard P2 Appropriate Policies and Procedures in *The Student Evaluation Standards (2001).*

> Student evaluations must be carried out consistently, equitably, and legally. Clearly written purposes, criteria, and procedures outlined in public policy statements and guidelines increase the likelihood that performance expectations for students will be understood, that a uniform standard of judgment will be applied, that evaluations will be fair, and that evaluation results will be trusted and used. Written policies provide students, parents/guardians, and other legitimate users with an opportunity to review and discuss the purposes of student evaluations and to ensure the protection of both teacher and student rights. Protecting these rights can promote trust in and support of the evaluation process and help to minimize the likelihood of litigation.
> Student evaluations conducted without a written policy and accompanying procedures run the risk of being poorly performed; variable across teachers, courses, and time; and susceptible to misuse and abuse, with no opportunity for review or recourse.

The original list included approximately 60 items that were either keyed to standard P2 or were tagged as policy related. These items were further categorized as school-based or teacher-based, reviewed to remove redundancies, and reworded in some cases to improve clarity. When an item addressed two or more different issues or concepts, I split the item into multiple items. In some instances an item was viewed as essential at both the school and teacher levels. In those cases the item was retained in both categories.

Once again the total list of items comprised an inordinately long checklist of approximately 50 checkpoints. Additionally, from practical and instructional viewpoints it seemed more appropriate to administer the school-based and teacher-based policies separately. For example, it seems appropriate to focus on teachers' student evaluation policies only when

school-based policies are in place. As a result I cut the checklist in two. This created two shorter lists, each more clearly focused, and removed the checklist redundancies that occurred from items that were pertinent at both the school and teacher levels.

☒   **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*

The two checklists have undergone several iterations. Initially, the checklists were presented unordered. After review (step 6 below) the checklists were divided into two parts. Here the division resulted partly as a means to simplify wording of the checkpoints (a header stem phrase is used for each section). More importantly, the items focus on two separate topics, what should be contained in the policy statements and conditions that the policies should meet. This organization is provided to help teachers and others organize their review of pertinent policies and better apply resulting findings.

☒   **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*

The checklists were prepared (in uncategorized form) and distributed for review in three ways. First, a draft was taken to a conference on classroom assessment and presented to teachers, administrators, and others for their review and comment. Second, a slightly modified form was given to a graduate student, Wendy Tackett, who both reviewed it and presented it to a collection of K-12 teachers for their review and comment. Third ,the checklist was reviewed by participants in The Evaluation Center's summer institute.

These reviews provided feedback regarding relevance (e.g., nearly all who reviewed the materials indicated their schools did not have comprehensive written policies on student evaluations), suggested modifications to the wording of several items, and relayed concerns related to the length (number of items and time required to complete the form). There were no suggestions to create item categories or to add items to cover additional topics.

☒   **7. *Revise the checklist content***

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

Suggested wording changes were addressed and several items were modified. The questions regarding the checklists' length and associated time to complete them led to related reductionist steps. First, Wendy Tackett, drafted much-abbreviated versions of ten and six items. That reduction process significantly simplified the forms, but skipped over or "lumped" together many items deemed essential to sound policies. Because each instrument can be read and completed in less than ten minutes and the two instruments are not intended to be used simultaneously, the concern for reducing their length seemed overstated. Also, significantly reducing the length would require ignoring Scriven's charge that "it is essential that it be complete." I decided against sharply reducing the content of the checklists.

Reviewer input did lead me to modify their formats to create their current form with two categories in each form. The two categories draw attention to and assist in focusing on separate but important aspects of the standards: (a) the content that should be included in the policies at the respective school and teacher levels and (b) the conditions that should be met in preparing and using the respective policies. Those changes break the instrument into smaller, easier to use parts, and hopefully reduce the apparent length of the instrument.

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

- *Determine with potential users whether category and/or total scores are needed or desired*
- *Determine with users what needs exist regarding differential weighting of categories and/or individual checkpoints*
- *Determine with users any checkpoints or categories of checkpoints that must be passed for a satisfactory score on the overall checklist*
- *Determine with users what needs exist regarding profiling of checklist results*
- *Format the checklist based on the above determinations*

I want the checklist to serve a metaevaluation role, assisting educators to determine whether their school and teacher student evaluation policies should be created or updated and improved. For those reasons I believe the checklist should provide "boundaries" for decision making. For example, "Is it okay to leave our policies alone for another year or so?" To encourage users to both think about such matters and facilitate discussion regarding such decisions, I arbitrarily constructed decision boundaries (cut scores) for each form. These boundaries now must be subjected to review.

☐ **9. *Evaluate the checklist***

- *Obtain reviews of the checklist from intended users and relevant experts*
- *Engage intended users to field-test the checklist*
- *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*

This evaluative step remains to be completed. My own sense of evaluative needs centers on two points: (a) completeness and (b) determining the viability of the current scoring system. This CDC checkpoint suggests additional points to be covered.

My interest in matters of completeness regard whether the forms overlook essential criteria and whether currently listed items can be merged together or removed without loss to the accuracy and utility of the checklist. Scoring system issues are more extensive and concern such things as Should the checklist items be totaled? Should cut scores be used? If so, should there be subtotals for the two categories? What total should be prescribed for the designated actions? Should actions other than those currently listed be included?

To address both my identified priorities and those noted above for this CDC checkpoint, I am engaging in several activities. First, a brief form is included in the presentation handout. I encourage you to review the checklist, complete the form, and return it to me in the provided envelope. Second, I encourage you to try one or both of the checklists in your own educational setting and provide feedback regarding its applicability. A second evaluation form is provided for that purpose. Third, persons who teach or work in K-12 school settings and participated in the Joint Committee's field trials for *The Student Evaluation Standards* are being invited to field trial the checklists and provide their feedback based on those trials.

☐ **10. *Finalize the checklist***

- *Systematically consider and address the review and field-test findings*
- *Print the finalized checklist*

I expect to complete the field trials for this checklist by the end of January 2002. At that point in time, the forms will either be finalized or revised for further field-testing.

☐ **11. *Apply and disseminate the checklist***

- *Apply the checklist to its intended use*

- *Make the checklist available via such means as journals, professional papers, web pages, etc.*
- *Invite users to provide feedback to the developer*

Once the checklist is completed, it will be published on The Evaluation Center's Checklist Web site (http://www.wmich.edu/evalctr/checklists/). Visitors to the Web site will be encouraged to use and report back on the viability of the form.

☐ **12. *Periodically review and revise the checklist***

- *Use all available feedback to review and improve the checklist at appropriate intervals*

Closing Thoughts.

I started down the path to develop this checklist as a means to help teachers and others apply *The Student Evaluation Standards*. While the *Standards* draft has received quite positive reviews, a consistently voiced concern is with the length (size) of the book. Teachers, other reviewers, and field trial respondents have suggested the importance of shorter, quicker and more easily used means to encourage use of the standards. We believe that the standards book should not be shortened, yet we also believe that full-scale attention to the guidelines and issues raised in the *Standards* book requires presentation of condensed information. These checklists are a step in that direction. Whether we are on the right track should become clear as steps 9 to 12 of the CDC are completed.

Finally, I thank both Drs. Scriven and Stufflebeam for their respective papers. I find both quite helpful to this checklist development process. This is the first time that I have applied the CDC to development of a checklist. In previous development of checklists I certainly have conducted many of the actions called for by the CDC. Yet, I find the checklist helps not just in an organizational sense but to encourage me to take all the steps necessary to establish a sound checklist. Similarly, I find Scriven's opening paper on the logic and methods for checklist to be a strong beginning point. Certainly, his characterization of the criteria of merit checklist helped me in determining which type of checklist best serves my needs. Also, as I decided against significantly shortening these two checklists, I found it comforting that he advocates listing all criteria of merit rather than a sampling of key issues. Both Stufflebeam and Scriven encourage use of their work and feedback for improvement. Please join me and others in the evaluation community in using these checklist development tools and providing feedback to increase their usefulness.

References

Brandt, R. (1989). On misuse of testing: A conversation with George Madaus. *Educational Leadership, 46*(7), 26-29.

Crooks, T., Kane, M., & Cohen, A. (1996). Threats to the valid use of assessments. *Assessment in Education, 3*(5), 265-285.

Gallagher, J. (1995). Education of gifted students—a civil rights issue. *Phi Delta Kappan, 76*(5), 408-410.

Gullickson, A. (1993). Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs, and challenges. In J. Conoley (Ed.), *Teacher training in measurement and assessment skills*. Lincoln, NE: Buros Institute of Mental Measurements.

Hills, J. R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan, 72*(7), 540- 45.

Joint Committee on Standards for Educational Evaluation. (2001). *The student evaluation standards.* (Manuscript in preparation).

Kelley, E. (1989). *Standards for grading and marking student performance*. Kalamazoo, MI: Joint Committee on Standards for Educational Evaluation.

Mayo, S. T. (1967). *Pre-service preparation of teachers in educational measurement* (Contract No. OE 4-10-011). Chicago, IL: Loyola University.

Sanders, J., & Vogel, S. (1993). The development of standards for teacher competence in educational assessment of students. In S. Wise & J. Conoley (Eds.), *Teacher training in measurement and assessment skills*. Lincoln, NE: Buros Institute of Mental Measurements.

Scriven, M. (2000). The logic and methodology of checklists. (http://www.wmich.edu/evalctr/checklists/)

Shafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57-63.

PRELIMINARY DRAFT

Shepard, L. (1989). Why we need better assessments. *Educational Leadership, 46*(7), 4-9.

Stiggins, R.J. (1991). Assessment literacy. *Phi Delta Kappan, 72*(7), 534-539.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: State University of New York Press.

Stufflebeam D. (2000). Checklists development checklist. (http://www.wmich.edu/evalctr/checklists/)

Teeter, A. (1995). Learning about teaching. *Phi Delta Kappan, 76*(5), 360-364.

Wagner, T. (1995). What's school really for, anyway? And who should decide? *Phi Delta Kappan, 76*(5), 393-398.

Wiggins, G. (1994). Honesty and fairness: Toward better grading and reporting. In T. R. Guskey (Ed.), *Communicating student learning: Association for Supervision and Curriculum Development yearbook*. Alexandria, VA: ASCD.

Wolmut, P. (1988). *On matters of testing misinformation*. Paper presented at the SRA, Inc. Invitational Conference of Testing, Phoenix.