

# SCORING PERFORMANCE ASSESSMENTS: CHECKLISTS, RATING SCALES AND RUBRICS

(Supports PEPE Teacher Indicator 3.2)

Alabama Department of Education Dr. Ed Richardson, State Superintendent of Education

Consultants: Dr. Judith Boser, Dr. Russell French, Mr. Joel McCay

**Institute for Assessment and Evaluation** 

**University of Tennessee** 

Developed with funding from Alabama's Title II
Teacher Quality Enhancement Grant

# Alabama Professional Development Modules

# Welcome Message

Dear Alabama Teacher or Administrator:

The Alabama Department of Education is happy to make available to you this professional development module. Created as part of Alabama's Teacher Quality Enhancement Project, these modules are intended to help you augment your knowledge and skills in the PEPE competencies and indicators. These competencies and indicators are the descriptions of teaching performance that serve as the basis for the PEPE Teacher Evaluation System. As you know, this information is based on research in effective teaching; i.e., teaching that produces student learning and best practice in Alabama classrooms and classrooms throughout the country.

We hope that this module, and the others that you may study, will be an enjoyable approach to professional growth. More importantly, however, we hope that it will be valuable to you and, ultimately, to those you teach.

Sincerely,

Ed Richardson

State Superintendent of Education

El Riberdon

# A. Introduction to Scoring Performance Assessments

This module has the following parts:

- A. introduction to scoring for performance assessments
- B. an information section on scoring for performance assessments
  - checklists
  - rating scales
  - rubrics
- C. a self-quiz titled "Check Your Knowledge"
- D. practice activity
- E. classroom application
- F. references and resources

To complete this module you will need pencil or pen and paper. It will also be helpful to have access to your gradebook for the previous year or a yearly plan that indicates the assessments used during the course of the year.

Alabama Professional Development Modules Scoring Performance Assessments: Checklists, Rating Scales and Rubrics

# **B. Scoring Performance Assessments**

Scoring is not always as simple and straightforward as counting the number of words spelled correctly on a spelling test. In performance assessments, students generate their own responses. Performance assessments frequently result in a product that is created or constructed by the students. Evidence of learning is demonstrated through such products as posters, essays, projects, models, research papers, and portfolios.

Another aspect of the performance assessment that may be assessed is performance itself, as well as the product created. Sometimes performance of something is the product. Performances are not limited to dance, physical education, art, and music, but can also include speeches, demonstrations, and class presentations in any subject area. Performance might also involve the actions Assess
both
products and
performance

of the student while preparing the product. In some situations, we observe how a student behaves when the student is "performing," but in other circumstances we observe learning-related behaviors in the classroom setting; e.g., time on task, participation in class discussions, oral presentation behaviors, or collaboration skills.. Whether we are assessing the product, performance, or both, one important guideline is that the assessment be done systematically so all students are assessed on the same basis.

#### **Checklists**

The least complex form of scoring system we will discuss in this module is the checklist. Checklists are appropriate when the teacher is looking for the presence of specific elements in the product or performance, and all elements are generally weighted the same. The teacher,

Mark
elements
or
behaviors
present

working from the prepared checklist, marks or checks each element as it occurs. The score is the total number of items that are checked. There is no extra credit or higher value received for doing one of the elements multiple times, unless the checklist specifies the number of times something must be done. For example, if one requirement was that a story to be written should contain three paragraphs, that item on the checklist would be checked only when there are three paragraphs. Various gradations in quality are not recognized unless the checklist specifies that something be done

"correctly," "appropriately," "neatly" or in some other way that brings quality into question.

Example 1 shows some items from an observation checklist used for parent-teacher conferences to evaluate a teacher's behavior in that situation.

Ex. 1	The teacher:
	Welcomes the parent/guardian by name
	Provides adult seating for all adults
	Begins with a positive statement about the child
	Correctly explains the student's test scores
	Solicits parents' questions
	Responds to parents' questions
	Summarizes the conference
-	ble 2, below, applies to basic speech skills students might be expected to demonstrate in g a three-minute presentation to the class.
Ex. 2	The student:
	maintains eye contact with the audience
	speaks loudly enough to be heard in all parts of the room
	enunciates clearly
	stands up straight (does not shift from foot to foot)
	does not go over the allotted time
	has notes
	uses notes sparingly
If stude	ents are assigned to construct a fire-prevention poster, the checklist below might be used:
Ex. 3	In the poster,
	Words are spelled correctly
	Lettering is evenly spaced
	No extraneous markings are on the poster
	Relevant drawings/figures/graphs illustrate the message
	Color is used

#### **Scoring for Checklists**

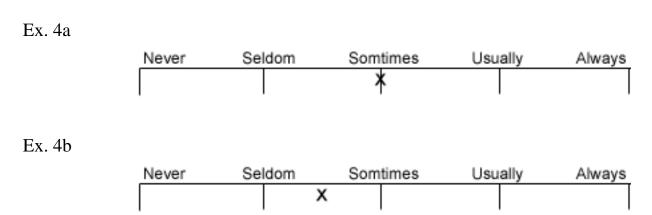
To determine a score for a checklist assessment, the number (or percentage) of items checked serves as the score. However, meaning must be attached to that score. What is the minimum score that would be considered adequate or passing? If letter grades are awarded, what score must be earned to receive an "A" on the assessment?

## **Rating Scales**

# Rating scales attach judgements about quality

The second form of assessment is the rating scale. Rating scales offer ways to attach quality to various elements of the process or product. For example, it is not uncommon for someone to ask us to rate something "from 1 to 10" with 10 being the highest score.

**Graphic rating scales** let you mark a point on a line or continuum to indicate your rating, as shown in the example below.



The second possible problem arises from the differences in perceptions of individuals as to the meaning of "sometimes" and "seldom" or any of the terms used above. The person doing the ratings is making a judgment based on his/her perception of the meanings of the terms in the rating scale. While I might think a student "sometimes" used correct grammar, another person might consider the same student's presentation as using correct grammar "seldom." We both observed the same behavior, yet we attached a different value to it.

**Numeric Rating Scales** are used sometimes in an effort to quantify results. We might ask the rater/observer/teacher to circle a number "1" to indicate that something never occurs, "2" if it happens seldom, "3" for sometimes, "4" for usually, and "5" for always. Note that here, as with the 1-10 scale, we have assigned the highest value to what we assume would be the desired response. This makes our ratings easier to interpret and makes the rating scale consistent with people's notions that better performance results in a higher score. Example 5a shows a rating scale with all the numbers defined, while Example 5b defines (or anchors) only the endpoints of the rating scale.

Ex. 5a

How often does the student	Never	Seldom	Sometimes	Usually	Always
Turn in lessons on time	1	2	3	4	5
Ex. 5b					
The student's essay	Never				Always
Uses correct capitalization	1	2	3	4	5

Numeric rating scales can be constructed without defining or giving a description of the meaning for every rating point. Example 5c also defines or anchors only the endpoints of the rating scale. In Examples 5a and 5b, we are looking at relative frequency of occurrence, from "not at all" to consistent occurrence. In Example 5c, we have a bipolar scale in which the "3" rating would represent the average expected level of performance. Performance that exceeds that would receive higher ratings, while sub-par performance would receive lower ratings.

Ex. 5c

The student's essay	Very Poor				Very Good
Organization	1	2	3	4	5
Appropriate capitalization	1	2	3	4	5

Numeric rating scales are particularly convenient when we are rating a number of qualities or behaviors on the same basis (in the case above, frequency of occurrence). The use of numeric rating scales overcomes some of the problems associated with graphic rating scales.

**Descriptive Rating Scales** are those in which each rating level is defined, often in detail and is not necessarily assigned a point value. Having good descriptions for rating levels alleviates some of the problems identified for graphic scales and does not force a teacher to quantify performance, if that is not appropriate.

#### **Scoring for Rating Scales**

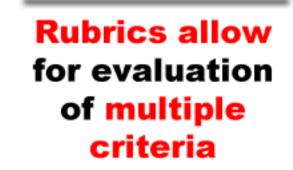
If there are several rating scales applied to a product or performance, as is commonly found, we can add the ratings to develop a total score. This procedure assumes that the difference between each pair of adjacent points on the scale is equal in value. In other words, the difference between "Seldom" and "Sometimes" is the same value as the distance between "Sometimes" and "Usually" or between "Usually" and "Always." This procedure also gives equal value to each criterion or item that is being rated (unless provisions are made to vary the weighting of the various criteria). As with checklists, there is still a need to determine the meaning of the score.

## **Rubrics**

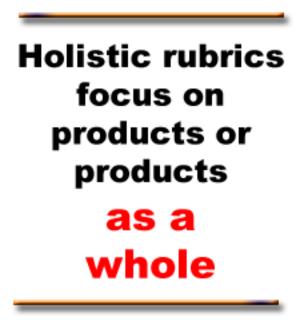
A rubric is a means of scoring a performance assessment wherein multiple criteria are being assessed and quality of performance or product is important. There are two main types of rubrics - holistic and analytic, although Nitko (2001) includes a third type, the annotated, holistic rubric.

In constructing a rubric, first determine that a rubric is the most appropriate way of scoring

the performance and/or product. A rubric is probably a good choice if there are multiple aspects of the product or process to be considered, if paper and pencil testing will not serve the purpose, and/or checklists and rating scales do not provide the breadth of assessment you need. Rubrics are especially appropriate for complex learning tasks or for those tasks and activities that integrate content from more than one content areas. The decision about the type of scoring needs to be made at the time the assignment is given to the students so that students can be given the scoring guidelines at the same



time they receive the assignment.



are not available.

A holistic rubric is most often used with writing and other creative or artistic endeavors. It focuses on the product or process as a whole rather than breaking it into separate categories (parts) and assessing each skill or criterion independently. A holistic rubric is a descriptive rating scale in which one score is awarded, with multiple indicators identified at each score level. For example, the statewide writing assessment in Tennessee uses a holistic rubric (shown below as it appears on the state website) to evaluate the student writing performance. This particular rubric, presented below, has six levels. Each level of performance is described or defined by several criteria. The resulting score (from 1 to 6) provides limited direction for improvement since performance on specific criteria

## **Tennessee Writing Assessment Scoring Rubric**

Scores from 6 to 1 reflect the range of excellence in the papers written in response to the assignment. Anchor and sample papers will illustrate both the limits and the diversity within each score level. Since each writing assignment places unique demands on students, readers score the papers primarily in reference to the standards that emerge from the sample papers. The following broad categories define the score ranges for the writing assignment and for the range of skills among students responding.

#### SCORE OF 6

A 6 paper is OUTSTANDING. It demonstrates a high degree of proficiency in response to the assignment but may have a few minor errors.

An essay in this category:

- \* is well organized and coherently developed
- \* clearly explains or illustrates key ideas

- \* demonstrates syntactic variety
- \* clearly displays facility in the use of language
- \* is generally free from errors in mechanics, usage, and sentence structure

#### **SCORE OF 5**

An essay in this category:

A 5 paper is STRONG. It demonstrates clear proficiency in response to the assignment and may have minor errors.

- \* is generally well organized and coherently developed
- \* explains or illustrates key ideas
- \* demonstrates some syntactic variety
- \* displays facility in the use of language
- \* is generally free from errors in mechanics, usage, and sentence structure

#### **SCORE OF 4**

A 4 paper is COMPETENT. It demonstrates proficiency in response to the assignment.

An essay in this category:

- \* is adequately organized and developed
- \* explains or illustrates some of the key ideas
- \* demonstrates adequate facility in the use of language
- \* may display some errors in mechanics, usage, or sentence structure

#### **SCORE OF 3**

A 3 paper is LIMITED. It demonstrates some degree of proficiency in response to the assignment, but it is clearly flawed.

An essay in this category reveals one or more of the following weaknesses:

- \* inadequate organization or development
- \* inadequate explanation or illustration of key ideas
- \* limited or inappropriate word choice
- \* a pattern or accumulation of errors in mechanics, usage, or sentence structure

#### **SCORE OF 2**

A 2 paper is FLAWED. It demonstrates limited proficiency in response to the assignment.

An essay in this category reveals one or more of the following weaknesses:

- \* weak organization or very little development
- \* little or no relevant detail
- \* serious errors in mechanics, usage, sentence structure, or word choice

#### **SCORE OF 1**

A 1 paper is DEFICIENT. It demonstrates fundamental deficiencies in writing skills.

An essay in this category contains serious and persistent writing errors or is incoherent or is undeveloped.

#### SCORE OF 0

A 0 paper is reported accompanied by one of the following codes to indicate a paper that could not be scored for one of the following reasons:

- 0-A Blank or Refusal
- 0-B Insufficient to score or Illegible

0-C - Off Topic

0-D - Written predominantly in another language

The **annotated holistic** rubric is a hybrid that starts out as a holistic rating but with comments added by the scorer to show examples that support the ratings. In other words, you might use the holistic rubric for a writing sample shown above, but after assigning the score you would briefly point out one or two strengths and one or two weaknesses that helped you decide on that score. Comments should relate only to the basis for your decision.

An **analytic rubric** provides scores for various skills or parts of the product or process, then combines the scores to obtain an overall total. Scores can be either summed to find a total score or averaged across the categories or criteria. Because separate aspects of the product or process are rated individually, the scoring provides more information to the student and teacher about the specific strengths and weaknesses of the student's skills than does a holistic rubric.

To construct an analytic rubric



**Step 1. Determine the criteria to be evaluated.** Each criterion should be labeled, and authors caution against lengthy lists of criteria. Probably three to five criteria can produce the best results because students can focus on them. With a long list of criteria, the students may have difficulty addressing them all and become confused as to their relative importance. (The same principle is used when supervisors provide feedback to teachers. Typically, a supervisor focuses on only a few weaknesses at a time rather than overwhelming the teacher with data that lead to overanalysis and paralysis.

A simple form for creating an analytic scoring rubric for use in performance assessment is shown in Example 6. The criteria are listed in the column down the left side of the page.

Ex. 6a

		Levels	
Criteria	1	2	3

Let's say our students have a reading exercise in which they read, or listen to, a book about William Bentley (the snowflake photographer) and another about Galileo. After the reading is concluded, our students are to identify three ways in which the stories in these two books are similar, find examples and illustrate them. This may involve skills of reading comprehension, art or technology (if illustrations are done with computers), and self expression, as they write a brief description of what they have produced. We have identified three criteria: Reading Comprehension, Illustrating, and Composition, as shown in Example 6b.

Ex. 6b

	Levels		
Criteria	1	2	3
Comprehension			

Illustrating		
Composition		
Composition		



#### **Step 2. Determine the number of performance levels.**

In the example above, there are three levels, with point values of 1, 2, and 3. We could have labeled our levels with other point values (such as 1, 3, and 5) to allow for some "gray areas" and give us the opportunity to record a score of 4 for an area that is not quite a 5, but definitely above a 3. Usually going beyond four or five levels makes it difficult for the scorer (and for the students) to distinguish between levels with any degree of confidence. While some rubrics have six or more

levels, it is more common to find rubrics with from three to five levels

**Step 3. Define expectations, beginning with the best.** Now it is time to actually define or describe what is expected at each level. Begin with the highest level of expected performance (in this example, the Analyst), as is done in Example 6c. Remember that level also has the highest score.

Ex. 6c

	Levels			
Criteria	Reader 1	Interpreter 2	Analyst 3	
Comprehension			Shows under-standing of the texts and recognition of similarities in the stories	

Illustrating		Illustrations are recognizable, in proportion, and accurate representations
Composition		Expresses thoughts clearly with appropriate grammar and spelling

After the highest category is described, the other levels are described. (See Example 6d.)

## Ex. 6d

Criteria	Reader 1	Interpreter 2	Analyst 3
Comprehension	Fails to under- stand the vocabulary in the books	Knows the vocabulary, but cannot recognize similarities	Shows under-standing of the texts and recognition of similarities in the stories
Illustrating	Illustration are unrecognizable or absent	Illustrations are poorly done	Illustrations are recognizable, in proportion, and accurate representations
Composition	Ideas are not clear	Ideas are expressed but with errors in grammar and/or spelling	Expresses thoughts clearly with appropriate grammar and spelling

#### **Student Involvement**

Teachers who involve their students in the creation of rubrics that will be used to score student work have found the student participation very helpful. The discussion leading to the final rubric also helps students clarify the expectations for their work.

#### Generic or task-specific rubrics

There are generic or generalizable rubrics, ones that can be used for all tasks of a specific type (such as public speaking, writing). Others are specific to a particular task. Repeated use of a generic rubric, such as one for writing, encourages students to improve from one assignment to the next because the objectives or criteria are the same. In some cases, the assignment or task to be evaluated may not be part of a series of similar tasks, or it may have one or more unique objectives. Students may do a project to demonstrate living conditions in the 1800s in America, but there may not be similar projects - or projects with similar objectives - during the year.

#### **Scoring**

In the analytic rubric, each criterion receives a separate score. Ordinarily the points for the criteria are summed for an overall score. They could also be averaged across the criteria for a final score. Using the writing exercise and rubric presented in example 6d, suppose a student's work demonstrated comprehension at the "3" level; illustrating was scored "2," and composition was "2." If the teacher summed the scores, this student would have received seven of a possible nine points. If the teacher averaged the scores, the student's score would have been 77.78

#### **Sources**

There are many Internet sites that provide a variety of ready-to-use rubrics or a basis from which you can create and print your own checklists and rubrics. While it is easier to use something that already exists, do so only if it is compatible with your objectives and learning activities. A few of these sites are listed in the References and Resources section.

### After scoring, what then?

No matter which type of scoring system we use, it is important to first determine what it is that we want students to demonstrate? Our instructional objectives should direct us toward the method of assessment, and this will guide our choice of scoring system.

While the scoring system we choose may yield a composite number, we must still attach a value to that number. If the highest score on an art project is 85, does that represent an "A" or is it a "C" based on our school district's grading system? If the 85 is the highest score, is there some area in which we have not provided sufficient instruction to the students? Or is there, perhaps, an area in which they did not expect to be evaluated and we neglected to address in class?

We can use our checklists, rating scales, and rubrics to analyze the overall performance of our class and, in effect, our instruction, much like we use the class mean, median, range, item difficulty, and item discrimination with written tests. We use the score earned by each student to calculate a class mean, median, and range.

#### **Analyzing Assessment Results**

We can apply some of the same concepts to the analysis of the results of our performance assessments that we use in analyzing tests When using a checklist, it is possible to determine what percentage of the students were checked or got credit for item 1, item 2, etc. Is there a pattern in that only a small percentage of students got credit on one or more items? If so, we may need to examine our instruction and our directions for the assessment. When we do this, we are using a procedure similar to determining item difficulty.

When rating scales are used, we can calculate the percentage of students receiving each rating for each item. If we find areas in which the class uniformly had low ratings, this has implications for us as educators.

When using rubrics, the process for analyzing assessment results is similar to that for a rating scale except that the score for each criterion is analyzed independently. On a written assignment, if we have criteria that relate to the content, organization, and language mechanics, is there an area of weakness for the class as a whole or do the scores spread across the rubric for all criteria?

This concludes the information section on Scoring Performance Assessments

Go to the next section to check your knowledge.

#### Alabama Professional Development Modules

Scoring Performance Assessments: Checklists, Rating Scales and Rubrics

# C. Check Your Knowledge

DIRECTIONS: On a sheet of paper, number from 1 to 7 down the left margin. Read each statement below and decide whether it is true or false. On your paper, write "true" or "false" beside the number corresponding to that statement.

True	False	1.	Checklists, rating scales and rubrics can be used interchangeably for the same purposes.
True	False	2.	Students are entitled to know before completing performance assessments how their efforts will be scored.
True	False	3.	Rubrics can be used effectively more than once with the same class.
True	False	4.	Creative activities are more likely to be scored with a holistic rubric than an analytic rubric.
True	False	5.	It is important to include in the rubric all possible criteria that can be evaluated in a single project or performance.
True	False	6.	Involving students in developing scoring rubrics is a waste of class time according to this module.
True	False	7.	A performance assessment can be evaluated on the basis of the product, the process used to produce it, or both.

Check your answers on the next page.

# C. Check Your Knowledge - Answers

Answers are shown in bold with asterisks.

True	False*	1.	Checklists, rating scales and rubrics can be used interchangeably for the same purposes.
		whil	cklists indicate only presence of an attribute, e quality is measured by rating scales, and tiple attributes are addressed with a rubric.
True*	False	2.	Students are entitled to know before completing performance assessments how their efforts will be scored.
True*	False	3.	Rubrics can be used effectively more than once with the same class.
True*	False	4.	Creative activities are more likely to be scored with a holistic rubric than an analytic rubric.
True	False*	5.	It is important to include in the rubric all possible criteria that can be evaluated in a single project or performance.
		_	number of criteria should be limited so students focus on them.
True	False*	6.	Involving students in developing scoring rubrics is a waste of class time according to this module.
			en students help develop the scoring rubric, they lop a better understanding of what is expected

of them. Discussion clarifies the expectations.

True\* False

7. A performance assessment can be evaluated on the basis of the product, the process used to produce it, or both.

If you missed two or more questions, you may need to review Section B of the Module.

If you missed only one or none of the questions, proceed to Section D.

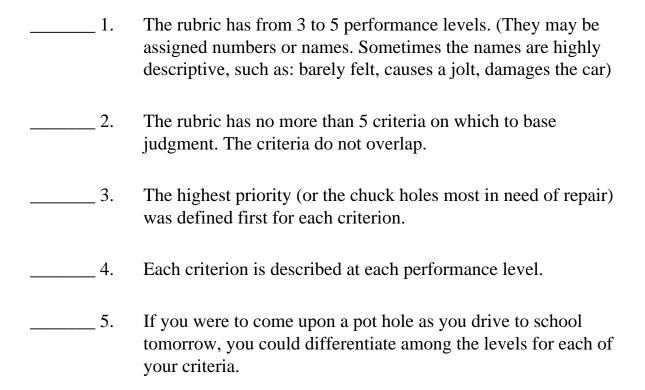
# D. Practice Activity

All of us are probably familiar with pot holes or chuck holes, those places in the road or street where a depression (sometimes large, sometimes small) has been created because some of the paving material is gone. The state highway department or local government is responsible for patching or fixing these holes. Have you ever wondered how they decide which ones to patch first?

Construct an analytic rubric that could be used by the highway department or by citizen volunteers to set the priorities for road repairs. After you have created your rubric, see the following section to evaluate it.

#### **Practice Activity Checklist**

Use this checklist to evaluate your rubric.



### Proceed to the classroom application in Section E.

# E. Classroom Application

Consult your grade book from last year or from the first part of the current year.

- 1. Make a list of the types of assessments you used in addition to tests and quizzes.
- 2. Indicate the type of scoring you used for each assessment from the headings of
  - a. checklist
  - b. rating scale
  - c. analytic rubric
  - d. holistic rubric
- 3. What scoring method or methods have you used?
- 4. Which method do you use most often?
- 5. Do you tend to use a general scoring system (the same checklist, rating scale or rubric) throughout the year or for multiple assessments, or different scoring for each task?
- 6. Is there an assessment for which you think a more complex scoring system would better suit your needs?
- 7. Select one assessment that you have used previously and plan to use again. Carefully examine the scoring system and make any changes in it that you think would improve the assessment. Discuss the task and your revised scoring system with your mentor, evaluator, or a colleague.
- 8. When you make the assignment, provide the students with a copy of the scoring system. Discuss the scoring system with them, giving examples of acceptable work or work at the various levels of the rubric, if available.
- 9. After student work has been scored, check to see how your students performed on each item on the checklist, level of the rating scale, or level of each criterion of the rubric. Is there a pattern that indicates a particular strength in your instruction? Is there a pattern that indicates an area in which further instruction is needed?
- 10. Discuss the rubric with the students again, if there appear to be problems with students' understanding of what is expected of them. If indicated by the discussion, revise the rubric before using it again.

This concludes the module on Scoring Performance Assessments: Checklists, Rating Scales and Rubrics. If you have questions or comments, please direct them to your mentor teacher or evaluator. That person can direct you to other resources if necessary.

Scoring Performance Assessments: Checklists, Rating Scales and Rubrics

## F. References and Resources

#### References

Airasian, Peter W. (1994). Classroom assessment. New York: McGraw-Hill, Inc.

Brookhart, Susan M. (1999). <u>The Art and science of classroom assessment: The missing part of pedagogy</u>. ASHE-ERIC Higher Education Report (Vol. 27, No. 1). Washington, DCP The George Washington University, Graduate School of Education and Human Development

McMillan, James H. (1997). <u>Classroom assessment: Principles and practice for effective</u> instruction. Boston: Allyn and Bacon

Nitko, Anthony J. (2001). <u>Educational assessment of students</u>. Upper Saddle River, NJ: Merrill Prentice Hall

Popham, W. James. (2002). <u>Classroom assessment: What teachers need to know</u>. Boston: Allyn and Bacon

Stiggins, Richard J. (2001). <u>Student-involved classroom assessment</u>. Upper Saddle River, NJ: Merrill Prentice Hall

Selected Web Sources for Rubrics and Checklists and On-line sites where you can design your own.

If the website URL or address has changed, use your search engine and search by subject or other keyword. Other state departments of education also provide good resources.

## http://4teachers.org/projectbased/

Age-appropriate, customizable project checklists for written reports, multimedia projects, oral presentations, and science projects.

http://www.odyssey.on.ca/~elaine.coxon/rubrics.htm - or http://www.quadro.net/~ecoxon/

The Ontario Schools, has examples of rubrics and links to rubrics for all or almost all content areas.

#### http://www.pals.sri.com/

Performance Assessment Links in Science. Performance tasks keyed to standards, include rubrics to evaluate student work.

#### http://rubistar.4teachers.org/rubric.php3?id=21&rubric=25

Another site where you can build your own rubric

#### http://www.uhslibrary.com/webquest.htm

Alaska Department of Education & Early Childhood Development Sample Assessment Tools and links to other sites