

AACC RESOURCE RATING GLOSSARY

This glossary supports the *AACC Comprehensive Content Centers Resource Review Form*, providing in-depth descriptions of the 1) types of resources evaluated, 2) prescreening criteria, 3) screening criteria, and 4) rating methods.

RESOURCES

Knowledge

Knowledge: usually research related resources as found in a journal, research report, or similar document approved through a peer review process. Because of the potential wide spread use or recognition of knowledge resources, they should generally be among the highest quality resources on the AACC web site.

- Example: Morgan, P., & Ritter, S. (2002). An experimental study of the effects of Cognitive Tutor Algebra I on student knowledge and attitude. www.carnegielearning.com/web_docs/morgan_ritter_2002.pdf
- Summary: In the 2000/2001 school year, the Moore, OK Independent School District conducted a study aimed at understanding the effectiveness of the Cognitive Tutor Algebra I program on students in their junior high school system. To allow the most rigorous test of the effectiveness of this course, the study was conducted as a true experiment, with students randomly assigned to either the Cognitive Tutor Algebra I course or a traditional Algebra I course. To control for teacher effects, some teachers taught both traditional and Cognitive Tutor courses. Dependent measures included the ETS Algebra I end-of-course exam, course grades and a survey of attitudes towards mathematics.

Product or Tool

Product or tool includes software, training materials, or publications that proscribe a specific process to follow or method to employ. Ideally, most products or tools in this category will have been reviewed externally and evidence will be provided that the product or tool has improved learning or achieved its goals.

- **Example:** Data Use: Data Primer-Learning Point Associates. <http://www.ncrel.org/datause/dataprimer/>
- **Summary:** The Data Primer is an instructional website to help educators use data to support improved learning. The Data Primer is organized around four modules asking the following questions. Where are we? Where do we want to go? How fast are we moving and in what direction? Are we leaving anyone behind? Each section has an on-line tutorial, a practice section, and further resources.

Guidance

Guidance, generally a written document that provides expert advice to improve educational assessment or accountability, that is based on current theory or research; state of the art knowledge; and has a substantial evidence base. The following would generally be considered *guidance*: congressional testimony, policy, information, or issue briefs, field guides, fact sheets, criteria for good data use, criteria for high quality ELL assessments, or assessment models.

- **Example:** *The Academic Performance Index: EdSource Online.*
http://www.edsource.org/pub_edfct_api.cfm

- **Summary:** This online information brief describes California’s Academic Performance Index, including a brief overview, the relationship between API and NCLB’s adequate yearly progress, the differences between base and growth scores, and goals for school improvement.. EdSource is an independent, nonpartisan, not-for-profit organization whose mission is to clarify complex education issues and to promote thoughtful policy decisions about public school improvement.

Services

Services include resources such as workshops, conferences, or expert technical assistance that are not practical for a peer review process, but that nonetheless are likely to be useful in helping states reach NCLB requirements. Services should include an evaluation component and have evidence of high quality such as established reputation of the source or use of materials based on rigorous evidence.

Example: The Council of Chief State School Officers sponsors *The Annual National Conference of Large-Scale Assessment*. CCSSO is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. http://www.ccsso.org/projects/National_Conference_on_Large_Scale_Assessment/.

Summary: This conference is the largest education assessment-only annual conference in the United States. With many prominent assessment researchers and practitioners presenting their work, the topics in 2006 covered the following broad assessment topics during a four-day period.

1. Assessing special populations
2. Issues related to compliance such as AYP
3. Innovative approaches in student assessment
4. Technical issues in assessment
5. Use of technology in developing, administering, scoring, and reporting assessments
6. Alignment issues
7. Formative assessment issues
8. NAEP
10. Assessment literacy
11. Operational steps in testing program
12. Assessment in content areas
13. Reporting/use of assessment data

PRE-SCREENING CRITERIA

Relevance to No Child Left Behind

Does the resource support the objectives of No Child Left Behind (NCLB)?

Descriptions of the Goals of No Child Left Behind (<http://www.ed.gov/nclb/landing.jhtml?src=pb#>)

Ensuring Student Learning

The principle of raising overall achievement and closing achievement gaps is paramount. Even as the benefits of an educated citizenry are enjoyed individually and nationally, states and local districts have the primary responsibility to ensure that all students are learning and that achievement gaps are closing.

Stronger Accountability

NCLB requires states to describe how they will close the achievement gap and make sure all students, including those who are disadvantaged, achieve academic proficiency. They must produce annual state and school district report cards that inform parents and communities about state and school progress. Schools that do not make progress must provide supplemental services, such as free tutoring or after-school assistance; take corrective actions; and, if still not making adequate yearly progress after five years, make dramatic changes to the way the school is run.

Proven Methods

NCLB puts special emphasis on determining what educational programs and practices have been proven effective through rigorous scientific research. Federal funding is targeted to support these programs and teaching methods that work to improve student learning and achievement.

Application to Resources

The above goals are represented on the AACC rating form in paragraph 1 as eight possible objectives. To be posted on the AACC web site, a resource should substantially meet a minimum of two of these objectives. It is not adequate for example, to simply mention an objective, such as Adequate Yearly Progress. Rather, the rater should feel confident that the resource, as a whole, would assist a user in meeting its adequate yearly progress goals. As is the case throughout this glossary, different types of resources may have different potential for relevance. A workshop for example, on formative assessment in middle school mathematics, may have more relevance to a narrow group of users than a national blue ribbon committee's accountability recommendations for closing the achievement gap written for policymakers.

Paragraph 2, content currency, is intended to remind the rater that educational resources are rapidly changing and frequently, although not always, improving. While a more recent resource may be more relevant, it does not guarantee relevancy, quality, or utility. A resource with an older development date may be equally or even more relevant than a resource with a newer development date, but raters are asked to substantiate relevancy of older resources with evidence.

Relevance to the Assessment and Accountability Comprehensive Center

Is the resource relevant to the AACC topics?

Below are the topics that pertain to the AACC, along with brief descriptions of each. For more information, please refer to the AACC website (<http://www.aacompcenter.org>).

Strand 1) Special Populations

The purpose of this strand is to provide resources to Regional Comprehensive Centers that will inform their work with states regarding the assessment and accountability of their special student populations (i.e., English language learners [ELLs], and students with disabilities, including students with the most significant cognitive

disabilities) with respect to the requirements of NCLB (i.e., Title I, Title III, 2 Percent Flexibility). Strand 1 work presents information on special student populations related to the following:

- Regulations and requirements per NCLB and other pertinent statutes and regulations
- Strategies for implementation/compliance
- Available resources
- Options and procedures for developing content standards
- Options and procedures for developing assessments
- Accommodations to limit bias and maximize reliability and validity
- Accountability systems and policies

Strand 2) Data Use or Formative Assessments

The purpose of this strand is to identify and share additional resources to Regional Comprehensive Centers that will expand state and local capacity to effectively use state, district, school, and classroom data that improve educational outcomes. In the first phase of work, Strand 2 drew from evidence-based research and well-known experts to identify a framework (ontology) that defined, clarified, and refined the domain of data use. Simultaneously Strand 2 collected the best available research and resources on data use, synthesized criteria for states and districts to use to evaluate available resources, and synthesized criteria for effective data use. Finally, Strand 2 derived and shared criteria for the effective use of formative assessment for instructional and program evaluation/review purposes. The Strand 2 focus in the second year was to provide RCC capacity-building opportunities for improved data use and formative assessment supporting NCLB goals. Year 3 will expand on those efforts, especially in the area of formative assessment.

Strand 3: State Assessment and Accountability Systems

The purpose of Strand 3 is to assist RCCs and states in implementing quality assessment and accountability systems consistent with NCLB requirements. The primary focus of this effort in the first year was to assist states in determining if they should apply for the federally sponsored pilot program in alternate AYP models. In year 2, Strand 3 shifted to a focus on assisting RCCs with the development and reporting of technical evidence to support state accountability systems. In year 3, a primary task is to assist the U.S. Department of Education in planning and facilitating the LEP Partnership¹, designed to assist state development and implementation of Title I ELL assessment methods.

Strand 4: Regional Content Center Initiatives

Increasingly, the AACC has received requests from regional centers for comprehensive, in-depth technical assistance to help address important NCLB-related assessment and accountability issues. The purpose of Strand 4 is to provide support where AACC capacity exists to meet specific regional initiatives. Examples of the type of work included in Strand 4 include but are not limited to: 1) a *one time big event*, e.g., support state/regional workshop, 2) *short term big issue*, e.g., support the resolution of important NCLB technical issues such as setting performance standards for a state assessment, or 3) *longer term big issue ("pilot")*, e.g., support the development and implementation of a district-wide or state-wide formative assessment plan.

IMPORTANT NOTES ABOUT RATINGS

Use of NA: Raters are encouraged to make ratings in all criteria whenever possible. For this purpose, raters may make *reasonable* inferences between the criterion and the resource. Consider for example the Cost-Effectiveness criterion. What may be a *cost-effective* new assessment program in a large school district with Title I funds may be cost-prohibitive in a small district without Title 1 funds. A reasonable inference for a rater would be to consider the cost of the new assessment program for a medium-size school district with some Title 1 funding. The rater might also be able to infer from available information if the new assessment system would replace an old system. Replacement would probably be more cost-effective than a add-on new program. Naturally, raters should

¹ The LEP Partnership is a U.S. Department of Education initiative to improve English language proficiency and core content assessments in reading and mathematics for limited English proficient (LEP) students.

consider using NA (Not Applicable or Not Assessed) when the criterion clearly does not apply to the resource, or in those cases when the rater's expertise is insufficient to make an informed rating.

Comments are extremely useful for every rating. Although it takes considerable time to write a comment for each criterion, comments help to validate the rating and the rater. While we encourage a comment for every rating, they are essential in those cases where a rater has given an NA.

Avoid overuse of the same numerical rating. Virtually all resources have substantial differences between the major scales (content quality, communications quality, utility, and evidence of effectiveness) and most resources will have differences within each scale. When raters give mostly the same rating, i.e., 1's, 2's, or 3's across the entire form or across an entire scale, it suggests that the rater did not do a thorough evaluation.

SCREENING CRITERIA DESCRIPTIONS

Content Quality

Goals: Does the resource clearly and briefly state its objectives?

A resource should introduce its topic area and convey its goals so that the audience understands the purpose of the resource. While not all resources will have an explicit introduction, the resource should provide a framework or context that communicates a need for the resource and the intended audience or audiences.

Standards: Was the resource developed using well-accepted literature, standards, research, or theory in the field?

Resources should make clear how they were developed so that users have high confidence in their probable effectiveness. *Knowledge* resources for example, should generally contain a review of the research literature on a given topic, and *Products/Tools*, *Guidance*, and *Services*, should also be based on scientific evidence or well-accepted theories and research. Low ranked resources may only have resource descriptions or individual endorsements. Resources that contain a bias review and external evaluation or review will generally receive a higher Standards rating than resources that do not.

Methods: Is the resource design based on high quality methods, i.e., randomized field trials, experimental studies, or other well-respected methods appropriate to the resource?

Are the methods behind the resource commonly considered "rigorous" (e.g. use of control groups, random assignment, experimental or quasi-experimental designs)? A *knowledge* resource that received a high rating, such as a research report, should generally itself use rigorous methods; whereas products, tools, guidance, and services that receive high ratings, should be based on work that uses rigorous methods. For example, a tool might receive a high rating based on an independent evaluation report. An upcoming conference might receive a high rating if the researchers presenting at the conference are known to use rigorous methods in their work.

Accuracy: Does the resource provide dependable information supporting its accuracy? A research report for example would likely include a literature review with a substantial number of references from peer-reviewed journals or sources known for accurate research and reporting. It would likely also discuss limitations of the research that might be due to the sample size lack or randomness in the treatment assignment. Is information presented without bias, for example, if particular point of view is presented in a knowledge brief, is an opposite point of view also discussed in a fair manner?

Plausibility: Are recommended steps, processes, or conclusions generally supported by other work or products in the field?

Are claims of effectiveness or impact realistic compared to similar resources? For example, are effect size increases for a specific program reasonable, or do achievement increases seem too good to be true? Can such

claims be independently verified? Is the resource supported by multiple studies or just one? Is the presentation of claims credible and include limitations of a study or resource? Or are claims overstated with bold proclamations?

Value: Does the resource add value to the field, or does it duplicate work or products from others?

While the term “value” can take on varying definitions for different people, does the resource contribute to the research or knowledge in accountability or assessment? Resources of high value, for example, may explore new subject matter not yet studied or clarify/critique current ones. Valuable tools, for example, may show how to apply what works in assessment or accountability so that the latest research-based information is used to support learning.

Communications Quality

Organization: Is the information well organized and presented in a logical sequence?

A resource should be presented in a format that is easy for the audience to follow. While not all resources will follow a prescribed format, specific resources should follow conventions and all resources should provide a framework that provides the reader with an overview or abstract of material presented, purpose, information and content that supports conclusions or a summation.

Language: How appropriate is the language (level of technical language) to the resource type, i.e., *knowledge, product or tool; guidance; or service?*

It is important that messages be conveyed effectively and clearly. Language should match the proficiency and background of the audience to maximize understanding and use. For example, a Fact Sheet for policy makers should be free of jargon often used by education researchers.

Length: Is the length of the resource or the time needed to use it appropriate for its goals and its audience? Consider breadth and depth of coverage.

Resources in different categories will vary substantially in length. A *knowledge* or *guidance* resource may exceed 30-50 pages and a *tool* may require a complete manual to assist the user. High ratings for length will generally include resources that are succinct, that is, they include the most important information and do not repeat it without a justifiable reason. Resources that provide several options to users, such as an abstract, an executive summary and a full report for a *knowledge* resource, will usually receive a higher rating than a resource that only provides a complete report.

Raters should consider the breadth and depth of the resource, recognizing that for some users breadth may be more important than depth and for other users, depth may be more important than breadth. Raters should therefore ask if the breadth and depth are appropriate to the resource type and to the intended audience. A policy brief on the purposes of accountability for example, will probably briefly cover many purposes, whereas a report on the same topic might cover fewer purposes but provide detailed tables and charts presenting data from a survey of school district administrators.

Visuals: Does the resource provide visual support for *important* information or *technical information*?

The use of charts, tables, and graphs is an effective means of supporting the resources’ main ideas. While not all resources will have visuals, such information is often helpful. Visuals should support the content and be clearly labeled. Most importantly, the visual representation should aid in the comprehension of the resource, not cause misunderstandings. Good visuals are usually brief and exclude extraneous information that might unnecessarily slow a user down.

Utility

Audience Relevance: Does the resource address an important or urgent need from Regional Comprehensive Centers or states?

The primary role of content centers is to inform regional centers and the states they serve, so raters should ask if these audiences really need this resource? Have they asked for it? If not, they are less likely to use the resource. It is possible that a state might use an AACC resource to create a new resource to inform school districts, schools and parents, but what evidence suggests that that will happen?

Timely: Can the resource be used or implemented in a timely period or will it take months or years to implement or use?

This criterion asks the rater to determine if the amount of time to implement a resource is worth the time. Raters should recognize that major educational changes most likely require a substantially longer amount of time before results occur and may require even more time for those results to be accurately evaluated.

Completeness: Are recommended steps, sequences and practices included?

Do stated findings and outcomes result from a logical progression of activities or practices that are clearly articulated? For example, are research and analysis methodologies articulated in knowledge resources and do products and tools, guidance and services detail and sequence steps necessary to achieve articulated results?

Cost Benefit: Does the expected improvement from this resource deserve the required investment of money and staff time? Specific questions to ask include: How much does this resource cost? What is the benefit, i.e., increased knowledge, increased learning? Will another program need to be eliminated in order to implement this one? If yes, will the expected improvement be worth the loss of the other program or product? If two programs run side by side, will they be compatible?

A knowledge brief might be inexpensive, for example, but its excessive length suggests to you that it will never be used. Low cost and low probable benefit should earn this resource a low rating. A new formative assessment program might be expensive to buy or develop, but the benefits in other schools have been very high in terms of increased student achievement. Further you know that the state will pay 50% of the cost. The rating should probably be high.

Supplemental Resources: Does the product provide supplemental resources, for example, references or a bibliography? If so, how useful are those resources?

Many resources include useful references or support materials. A research paper that receives a high rating on this criterion, for example, should include a substantial number of references from other well-known researchers on this topic. A two-day professional development program on assessing special needs students, should have supplemental resources such as a guidebook or handbook that will help participants use or share the new knowledge in their classroom, school, or district.

While a rater may not have immediate access to these supplemental resources, she or he may recognize authors of the resources or may be able to obtain added information by doing a simple Internet search. Consider as well the quality of the company or organization producing the supplemental resource.

Evidence of Effectiveness

Impact: What evidence supports that this resource has had or will have substantial impact (i.e., narrowing of achievement gaps, policy changes, improved achievement, etc)

What has been the impact of this resource? Have those outputs been positive? Were there any negative effects and if so, how major was the effect? How confident are you that the impact was accurately reported? Were claims

for a product, for example, actually met? An external evaluation of a product or tool from a credible source, for example, would likely have a higher impact rating than results from an internal evaluation.

As with all criteria, impact will differ based on resource type and purpose. In general, greater promises or claims made by an author or organization, or for the consequences of implementing that resource, require greater evidence of impact. An assessment or workshop, for example, that claims to reduce learning gaps for ELL students, needs substantial evidence to support its claim compared to a similar assessment or workshop, which claims only to expand users knowledge.

Generalizability: What evidence exists that this resource has been effective or will be effective in many different settings?

Resources with strong external evidence of working well in multiple and diverse settings should receive higher ratings than resources with demonstrated effectiveness in only one or a few schools, or with no evidence at all. A high rating would likely be given to a resource where an external evaluation found strong evidence of a program's positive impact on many schools in diverse geographic settings and in different types of schools, i.e., urban, rural, and suburban. As with other criteria, raters should evaluate this criterion based on the type of resource and the claims of its developer.

RATING METHOD

Rating Scale

The rating scale ranges from 0-3, including a Not Applicable or Not Assessed for the instances that a specific criterion does not apply to the reviewed resource or has not been rated by the reviewer. Below are descriptions for each of the rating levels:

- 0= Weak or not at all-- There is little or no evidence that the resource meets the specific criteria being reviewed.
- 1= Fair-- The resource displays some evidence of the specific criteria being reviewed, but this evidence is minimal.
- 2= Good-- The resource displays a fairly significant amount of the specific criteria being reviewed, but not overwhelmingly so.
- 3= Excellent-- The resource displays sizeable and highly dependable evidence of the criteria being reviewed.

Rating Calculations

Within the rating criteria there are four sections (Quality Development, Quality Communications, Utility, and Evidence of Effectiveness), each with specific questions that are to be rated on a scale of 0-3, if applicable. Reviewers are then asked to calculate the average ratings for each section and *the average of all four averages will constitute the resource's overall rating*. If a question is deemed "not applicable" or "not assessed" then the section average rating should reflect this by adjusting the denominator of total questions accordingly.

There is also a holistic rating question at the end of the rating criteria that asks the reviewer of their likelihood to recommend the resource to a colleague, on a scale of 1-5. While this rating will not factor into the overall rating of the resource, it provides additional information on whether the resource should or should not be posted on the AACC web site.

AACC Spotlight

Your recommendation to include a resource as a Spotlight feature is very important. In general, the intent is that after committee review and approval by the AACC director, new content will be featured for a one- to two-week period as a Spotlight. However, some material may be dated and not necessarily appropriate for a Spotlight. Or it may have been so prominently featured elsewhere that a Spotlight is not necessary. Consider these factors into your recommendation.

Be aware also that some materials may be so newsworthy, of such urgency, or of obvious U.S. Department of Education support, that a full committee review may not be necessary. In these cases, the communications director may request posting of such items as Spotlights with approval only from the AACC director. For example, new guidance from the U.S. Department of Education would be deemed to have U.S. Department of Education endorsement and to support state level NCLB goals, that it could be a Spotlight feature once approved by the AACC director.

Certification of Non-involvement

The Assessment and Accountability Comprehensive Center desires that all products listed on its website be unbiased and fairly rated. Reviewers who participate in the rating process have reviewed the AACC Certification of Non-involvement and stated that they have no financial interest in any resource and/or have not participated substantially in a rated resource's development.

Guidance examples - Invited presenters at a conference might participate in the review of a conference, however conveners of a conference will not. A reviewer may review a paper developed by his or her own organization only if the reviewer is not an author of the same paper and did not take part in any aspect of the paper's development.

All financial or personal involvement in any resource reviewed must be disclosed prior to a review. Any reviewer who believes that they should not participate in a review for any reason may rescue himself or herself without further discussion. The AACC director will be contacted to resolve specific questions of potential bias and if necessary may consult with U.S. Department of Education officials.

Acknowledgement

Our appreciation to the National High School Center for allowing us to adapt their rating form and glossary for AACC purposes.