

Assessment Critiques

Analysis of Sample 1: Everett Math Assessment and Instructional Guide

Intended Grade Levels: Grade 4–12; Grade 6 reviewed

Key 1: Assessment serves a clear and appropriate purpose. Did the author specify users and uses, and are these appropriate?

The authors clearly articulate the purposes for both the large-scale assessment and the classroom assessment ideas presented in this booklet. The goal is to “provide meaningful information back to teachers on how our students performance on this assessment, how those results can be used to better inform math instruction, and how to improve student performance on future assessments” (p. 4). The booklet provides students and teachers with samples of quality work that can be used instructionally to teach the very skills also being assessed.

Users and uses are focused. There aren’t too many. It is very clear how assessment of and for learning fit together.

Rating: We would give this assessment a “5” on the trait of clear and appropriate purposes.

Key 2: Assessment reflects valued achievement targets. Has the developer clearly specified the achievement targets to be reflected in the exercises? Do these represent important learning outcomes?

The authors have clearly specified the achievement target to be assessed and it is selective and easy to find. They note that the focus is math problem solving and include rubrics and samples of student work for defining what they mean. The target is clearly related to state content standards. The target is important and worth the time devoted to it. It is clear that the learning target came first and then instruction and assessments were designed to track progress toward and help students attain the targets.

Rating: We would give this assessment a “5” on the trait of clear and appropriate targets.

Key 3: Design. Does the selection of the method make sense given the goals and purposes? Is sampling appropriate to get a good estimate of student learning? Is there anything in the assessment that might lead to misleading results?

Choosing the Best Method. The assessment method matches the purpose and target. The authors are assessing reasoning, problem solving, and communication in math. These are all assessed well with extended written response and performance assessment. There is a wonderful alignment between the targets to be addressed, instruction, and the actual assessment. The link illustrates exactly how to use large-scale performance assessment materials in the classroom to build the very skills also being assessed. (5)

Writing Questions. The tasks that are provided as examples are clear. The problems include a statement of the problem, requirements of showing and explaining the solution, and the rubric used to evaluate responses. (5)

The rubric might be fine tuned. Although the four dimensions, or traits, represented in the rubric are fairly common, the detail provided to describe score levels is sometimes vague. For example a “5” in “communication” includes, “Explains thinking thoroughly.” A “4” is, “Explains thinking basically.” A “3” is, “Partially explains thinking.” A “2” is “Attempts to explain thinking.” This raises some questions. For example, what exactly is the difference between “partially explains thinking” and “attempts to explain thinking”? Also, when is explanation of thinking thorough enough to warrant a “5”? There are similar questions with many of the other descriptors in the rubric.

Sampling. This is the weakest element in the assessment as described. To decide on a good sample we need to consider the breadth of the target, the coverage of the task(s), and the stakes.

First the large-scale assessment. It seems that students are given two problems to solve. Since each problem is small in scope, this is unlikely to adequately cover either the sixth-grade content to which problem solving might be applied, nor the range of possible problem solving strategies. To provide a snapshot of problem-solving ability of the group, and to provide information to make instructional decisions about individual students, we need more.

Second, only having two sets of samples to illustrate quality could be bothersome. If students (or teachers) only see models of quality based on two problems, they are less likely to be able to generalize to a broad array of other problems. They might think that the way shown in the sample problems are the only acceptable way to solve a problem. This restricts students’ vision of the real target. Students should have the opportunity to critique lots of examples of problem solving that cover a variety of content and strategies. (3)

Sources of Bias. The authors discuss two sources of bias (although they are not identified as such): rater agreement and test preparation practices. They do not however, explain the reason for addressing these issues: to obtain the most accurate estimate of student achievement. It is also unknown whether students have an alternative to writing out their solutions (for those students who don't write well in English). (4/5)

Rating: We would give this assessment a “3/4” on the trait of design.

Key 4: Communication. Is it clear how this assessment helps communication with others about student achievement?

Although communication is not specifically addressed, the document meets many of the requirements of sound communication. The whole goal of the document is to build common understanding of the target on the part of teachers, students, and parents. The document reports results and explicitly describes many strategies to improve it. It would be nice to see something about descriptive feedback to students.

Rating: We would give this assessment a “4/5” on the trait of communication.

Key 5: Student Involvement. Is it clear how students are involved in the assessment as a way to help them understand achievement targets, practice hitting those targets, see themselves growing in their achievement, and communicate with others about their success as learners?

This is a strength of this document. The document helps teachers use the assessment as a tool for learning in the classroom as well as helping teachers to understand the math problem-solving learning outcome. The authors give many suggestions for using the assessment materials for instruction. They describe procedures to help students understand the achievement targets they are to hit, practice hitting the target, and communicate effectively with others about their success. There are student-friendly versions of the rubric. Student involvement is meaningful.

We couldn't find assistance with helping students “see themselves growing in their achievement over time,” but this could easily be added.

Rating: We would give this assessment a “5” on the trait of student involvement.

Overall Judgment

This assessment and associated materials and procedures has many strengths. The major things that could be fine tuned relate to sampling and the wording in the rubric. We would give it a “4/5.”

Analysis of Sample 2: Fish Tank

Intended Grade Levels: Grade 5

Key 1: Assessment serves a clear and appropriate purpose. Did the author specify users and uses, and are these appropriate?

The description does not specify users and uses. It is not clear why the assessment is being given. It is not clear how the assessment would inform future instruction.

Rating: As presented, we would give this assessment a “1” on the trait of purposes.

Key 2: Assessment reflects valued achievement targets. Has the developer clearly specified the achievement targets to be reflected in the exercises? Do these represent important learning outcomes?

This assessment has at least one strong feature relating to targets. The developer has provided test specifications. This helps to clarify the learning targets the developer intends for this assessment. The targets are stated, selective, and easy to find. Further there is a mix of targets and a conscious attempt to assess more than recall of facts.

We have several questions, however, relating to the importance and clarity of the targets: Are the knowledge targets worth the instructional time devoted to them? How do they relate to district/state content standards? Do they represent best thinking in the field? How do they relate to the teacher’s overall learning plan for the year? Does the blueprint offer a good testing plan, given the content of the lesson?

Another apparent strength is the plan to assess student knowledge of how to set up a fish tank before actually asking students to do it. This indicates that the developer thought through a sequence of instructional tasks and assessments and how they relate. There is, however, no attempt to assess the extent to which students actually did adequately set up a fish tank. This gets back to the clarity and importance of the targets being assessed.

It probably is not worth the time to do a performance assessment on how well students actually can set up an aquarium. Therefore, why is this task being done at all? Are there other important learning targets being addressed: group skills, for example? If there are other important targets, why aren't they listed?

Rating: Although the targets are clear, there is a question on importance. We would give it a “2” on a scale of 1–5, where 1 is low and 5 is high.

Key 3: Design. Does the selection of the method make sense given the goals and purposes? Is sampling appropriate to get a good estimate of student learning? Is there anything in the assessment that might lead to misleading results?

Choosing the Best Method. Target–method match is one of the strengths of this assessment. The focus is on knowledge and reasoning. Selected response and essay formats are good matches for this type of learning target. Further, there is a mix of methods. (5)

Sampling. This assessment does test the knowledge and reasoning proficiency of students on how to set up a fish tank. Further, the essay can tap evaluative reasoning. But the overall question here is, Do the items cover in the proportions suggested by the opening table of specifications? We think not. There is a disproportionate representation of knowledge items. (2)

Writing Questions. Many of the test questions don't adhere to standards of quality for selected response and essay formats. For example, in question 4, at least two responses are ruled out due to grammar. The answer to question 6 is given away because of the repeat of words from the question to the right answer. In question 8, the use of a negative makes the question harder to answer.

In the matching section, some options have more than one correct answer. Some options are confusing. For example “ready to go” probably refers to the gravel because the lesson text uses these very words. But it's hard to be sure; lots of other things on the left hand list also come “ready to go.” This list of options is very, very long. Who can sort through this much information effectively?

In question 10, the intended answer is 80 degrees, but this may be very unclear. Does the author mean for the student to write “80 degrees” in the two spaces?

In question 11, there are so many blanks that unexpected responses might be correct but entirely miss the point. For example, “After your fish tank has cured for more than

a week, add a few healthy fish for every 10–20 gallons of liquid.” These answers are not wrong but are not what the author is expecting.

In question 13, only one of these taps evaluative reasoning. The other calls for a regurgitation of material presented in the passage. (1)

Sources of Bias. No other sources of bias.

Rating: We would give this assessment a “2” on the trait of design. The overall quality of the test questions is negative.

Key 4: Communication. Is it clear how this assessment helps communication with others about student achievement?

The author has specified no plans for communication.

Rating: As presented, we would rate this assessment a “1” on the trait of communication.

Key 5: Student Involvement. Is it clear how students are involved in the assessment as a way to help them understand achievement targets, practice hitting those targets, see themselves growing in their achievement, and communicate with others about their success as learners?

The developer does not mention using the assessment materials or results in this way.

Rating: We would give this assessment a “1” on the trait of student involvement.

Overall Judgment

The weaknesses far outweigh the strengths in this assessment, but there are some strengths—a table of specifications, and good target-method match, for example. The biggest problems relate to importance of targets, the quality of the test questions themselves, the lack of a stated purpose, no provision for communication, and no student involvement. We’d give it a “1/2.”

Analysis of Sample 3: Culminating Project

Intended Grade Levels: Grades 8–9

Key 1: Assessment serves a clear and appropriate purpose. Did the author specify users and uses, and are these appropriate?

The author alludes to the purpose of the assessment—to document competence. But, there is no real statement of how the results will be used and who will use them. Will this result in a grade? A judgment of overall mastery? Is this a barrier exam—do students have to do well to progress to the next grade? Is this purely for information—teachers might use it to plan instruction and parents might use it to judge the progress of their children? We simply don't know.

Rating: We would give this one a “2” on clear and appropriate uses and uses, on a scale of 1–5, where 1 is low and 5 is high.

Key 2: Assessment reflects valued achievement targets. Has the developer clearly specified the achievement targets to be reflected in the exercises? Do these represent important learning outcomes?

The author has listed targets and they are easy to find. The targets seem to be important and worth the assessment time devoted to them. There is an appropriate mix of targets. There is some evidence of long-term thinking—the overall big picture of the skills, knowledge, and dispositions to be developed in students. But, these are far outweighed by several major problems:

- Targets are general and vague. Educators are likely to interpret the targets differently. For example, what does “good study habits” entail? Or “group discussion”?
- Everything is listed. Granted, this is a culminating activity, intended to assess overall student master's of the skills outlined in the social studies curriculum for grade 8 or 9, but, on first glance, it appears that it will be challenging to fit them all into a single assessment.

- There is no effort to relate targets to local content standards. We have inferred that this is so based on other statements in the assessment (“exit outcomes” is the title for the targets).
- The assessment itself does not clear up the nature of the targets. For example, the rubric provided only helps define a very small part of the targets listed.

Rating: We would give this one a “2” on clear and appropriate targets on a scale of 1–5, where 1 is low and 5 is high. The targets are important, but are not well defined. Additionally, there are so many listed that it is difficult to see how they can all be assessed in the context of a single assessment. Perhaps the task does require all these skills, but that is different from actually assessing them.

Key 3: Design. Does the selection of the method make sense given the goals and purposes? Is sampling appropriate to get a good estimate of student learning? Is there anything in the assessment that might lead to misleading results?

Choosing the Best Method. The complex nature of the targets require a complex assessment, as outlined in this assessment. So, some sort of performance assessment is in order. But, this is outweighed by several major problems:

- There is no stated rationale for the method used.
- There are lots of missed opportunities here. There might be occasions during the research and preparation for this exhibition during which many of the listed exit outcomes can be assessed. For example, if students are working together, the teacher might assess *respectful*, *working cooperatively*, *effective listening*, and *group discussion* during one or more work sessions. Or, students could continually self-assess their own progress. (But, where are the rubrics for these?)
- Only one small part of the targets is covered in the assessment. (1)

Writing Questions. It is only partially clear what each student is to do—the tasks are only partially explained. Since this is such a big task, there are lots of opportunities for a lack of clarity.

The rubric and criteria are the real problems here. Does the author really believe that all of the student effort put into the task, and all the “exit outcomes” listed, are adequately reflected in the rubric? (1)

Sampling. We are nervous about the sampling. The assessment aims at broad targets and the stakes are seemingly high. This requires careful sampling to ensure that inferences about student capabilities are accurate. Granted, the task is very complex and does require application of lots of skills. So, it does cover a lot of ground. But, it’s still one shot, sink or swim. We would feel better if we knew the students had practice opportunities before the “real” exhibition, and if they had opportunities to demonstrate prerequisite skills embedded into instruction all year. We might presume this is true, but we really don’t know.

There are lots of targets that aren’t assessed at all. How does this assessment, for example, get at the targets listed under “Essential Questions”? Is application of this knowledge an important part of the task? If so, where? And even if it is an important part of the task, where are the rubrics to judge application of this knowledge in the context of this task? Performance assessments require both tasks and criteria to be assessments. (1)

Sources of Bias. The author doesn’t address the issue of bias and distortion at all. In fact, the author doesn’t seem aware that it is important to consider such issues. Is this an equally appropriate task for a diverse group of students? Are the reading, writing, and presentation requirements equally appropriate? Might there be some features of the task that will mask students’ true ability? For example, the major products are a written report and an oral presentation. While those are legitimate targets in and of themselves, having to write and present such a long piece might mask student ability to read, understand, and reason. The opportunity to present research orally, 1:1 with the teacher, might be a more valid way to assess reading, understanding, and reasoning for some students. Student ability to write about their knowledge could be assessed separately. This would enable the teacher to better tease apart the strengths and needs of each student. (1)

Is the task developmentally appropriate for middle school students? Has the needed instruction, and learning, occurred that puts in place prerequisite skills? Or, is this too much to ask for middle school students?

Rating: We would give this assessment a “1” on the trait of design.

Key 4: Communication. Is it clear how this assessment helps communication with others about student achievement?

The author has not considered communication at all. There is a rubric for one of the products in the assessment, but other than to assess the quality of the research report, we do not know how it might or might not be used to communicate achievement. Further, the rubric provided is incomplete—it is hard to see how communicating with this rubric would help. Finally, the author does not provide a mechanism for reporting on any of the other targets listed.

Rating: We would give this assessment a “1” on the trait of communication.

Key 5: Student Involvement. Is it clear how students are involved in the assessment as a way to help them understand achievement targets, practice hitting those targets, see themselves growing in their achievement, and communicate with others about their success as learners?

There is no student involvement.

Rating: We would give this assessment a “1” on the trait of student involvement.

Overall Judgment

This assessment is weak. It would require a lot of work to make it usable. It would be better to look elsewhere. We would give it an overall rating of “1.”

Analysis of Sample 4: Emerson Essay Test

Intended Grade Levels: Grades 10–12

Key 1: Assessment serves a clear and appropriate purpose. Did the author specify users and uses, and are these appropriate?

The purpose is clearly stated—summary assessment. The users and uses are focused.

Rating: We would give it a “5” on the trait of clear purposes.

Key 2: Assessment reflects valued achievement targets. Has the developer clearly specified the achievement targets to be reflected in the exercises? Do these represent important learning outcomes?

The developer has stated that the targets are knowledge of Emerson and making inferences about Emerson's stand on current issues. These are selective, easy to find, clear, and a good mix. However, there is no statement as to their importance. How do these targets relate to local content standards? Why is knowledge of Emerson important?

Rating: This assessment is a balance of strengths and weaknesses. We would give it a "2/3" on a scale of 1–5, where 1 is low and 5 is high.

Key 3: Design. Does the selection of the method make sense given the goals and purposes? Is sampling appropriate to get a good estimate of student learning? Is there anything in the assessment that might lead to misleading results?

Choosing the Best Method. The assessment method matches purpose and target, this is a clever use of extended written response, and there is a good match between the targets, instruction, and assessment. (5)

Writing Questions. It is clear what students are to do; in fact, they've practiced it. The tasks match the targets and they are feasible. The lack of criteria for making inferences is problematic. (3)

Sampling. There are enough samples of student performance to get a stable estimate of the learning target, and the sample matches the breadth of the target and importance of results. (5)

Sources of Bias. Might some students be able to reason adequately but score low because of their writing skills? (4)

Rating: This assessment is stronger than weak on the trait of design. We would give it a "4."

Key 4: Communication. Is it clear how this assessment helps communication with others about student achievement?

The author does not directly discuss issues of communication, although the scoring and grading procedure is described and it is reasonable. Communication could use work on understandability and descriptiveness.

Rating: The assessment displays a balance of strengths and weaknesses—we would give it a “3.”

Key 5: Student Involvement. Is it clear how students are involved in the assessment as a way to help them understand achievement targets, practice hitting those targets, see themselves growing in their achievement, and communicate with others about their success as learners?

Not mentioned.

Rating: We’d give this assessment a “1” on student involvement.

Overall Judgment

This assessment is a balance of strengths and weaknesses— we would give it a “3.”

Analysis of Sample 5: Reading Rate Assessment

Intended Grade Levels: Grades 2–3

Key 1: Assessment serves a clear and appropriate purpose. Did the author specify users and uses, and are these appropriate?

Although the assessment developer did not specifically list users and uses, it is easy to figure them out. The intended purposes appear to be to (1) help the teacher continuously track student progress toward a well-defined target—a reading rate of 110 words a minute; and (2) keep parents informed of the progress of their children about their ability to read grade-appropriate material. In this case, ability to read is measured by reading rate. The users and uses are focused, clear, and appropriate, but they have to be inferred.

Rating: We would give this assessment a “4” on the trait of clear and appropriate users and uses, on a scale of 1–5, where 1 is low and 5 is high.

Key 2: Assessment reflects valued achievement targets. Has the developer clearly specified the achievement targets to be reflected in the exercises? Do these represent important learning outcomes?

The target is clearly stated as being reading rate. The developer even gives a specific target reading rate and the formula for determining rate—number of words read in a minute minus any words read incorrectly. It is stated, selective, and everyone would interpret the target the same. (5)

However, we have questions. Is reading rate the most important thing to assess in reading? Is it the only thing this teacher assesses in reading? Is it worth the time devoted to it? The developer’s assertion about the relationship between reading rate and comprehension is probably true, but one might read pretty fast in Spanish and not understand much. So, the developer may need to review her mix and relative emphasis of targets—there is no evidence of an overall plan to cover all important goals. (1)

There is no link to standards, but there is a reference to best thinking in the field. (3)

Rating: We would give this assessment a “2/3” on the trait of clear and appropriate targets, on a scale of 1–5, where 1 is low and 5 is high. Although this target is very clear, we have troubling questions about its importance (appropriateness).

Key 3: Design. Does the selection of the method make sense given the goals and purposes? Is sampling appropriate to get a good estimate of student learning? Is there anything in the assessment that might lead to misleading results?

Choosing the Best Method. It is clear that the developer picked the best procedure for assessing reading rate—performance assessment. If one wants to see how fast students can read, the best way to tell is to have them read something and time it. The developer has also provided a reasonable rationale for the method chosen. (5)

Writing Questions. The task is certainly clear, aligns with the target, and is feasible. Several questions arise, however. First, how did the developer determine that the selection to be read was at the second- (or third-) grade level? If the determination of level of

reading passage is off, then reading rate will be affected. If the selection is too easy, then reading rate might be too high; if the selection is too hard, then reading rate might be too low. We would like to ask the author how the books are selected.

The criteria used to determine success are clear. This assessment attempts to measure a fairly straightforward target, so long, extensive rubrics are not needed—simple target, simple rubric. So, the procedure for determining rate is certainly reasonable. However, we'd still like to do more research on the formula used to determine rate. Is total number of words read in one minute minus number of words read incorrectly the commonly accepted way to determine reading rate? (2)

Sampling. The reading rate activity described does not cover the domain of achievement implied by the target. It appears that the developer relies on a single one-minute sample of a single reading selection. Might the average rate over several minutes result in a more stable measure of rate? Might reading rate averaged across several different selections of various types result in a more valid measure of rate? This assessment, as stated, might be one good measure of rate, which, when combined with others, might provide a stable estimate of overall reading rate. Additionally, there is the lingering question about the role of reading rate in reading comprehension. If the real goal of the assessment is to provide an estimate of reading comprehension using reading rate as a surrogate measure, then the method and sample described is probably not enough to draw a conclusion about comprehension. (1)

Sources of Bias. Are the books selected appropriate for all students? Is knowledge of snow, for example required for understanding *Look Out Ronald Morgan*? If so, might students from the South Pacific be at a disadvantage even if their reading rates were the same as students who do know about snow? We're not saying that this is a problem, it is just that there is no evidence that the assessment developer has asked this question. A solution might be to have several books, equivalent in difficulty, that are matched to students. (3)

Rating: We would give this assessment a "2" on the trait of design. If we were to use this assessment, we would want to corroborate the developer's conclusion on the formula and level of books, and supplement the list of selections students can read.

Key 4: Communication. Is it clear how this assessment helps communication with others about student achievement?

The developer appears to have thought this out. The developer has made it very clear what the target is and has set up procedures for continuous reporting to parents. The overall idea is sound.

However, are the letters, as given, the best way to communicate with parents? Will they understand them? Would the tone of either letter put parents off?

Rating: We would give this assessment a “3” on the trait of communication.

Key 5: Student Involvement. Is it clear how students are involved in the assessment as a way to help them understand achievement targets, practice hitting those targets, see themselves growing in their achievement, and communicate with others about their success as learners?

The developer does not mention using the assessment materials or results in this way, there is no student involvement component and there is nothing to indicate that this is part of a bigger student involvement plan over this.

Rating: We would give this assessment a “1” on the trait of student involvement—the potential is there, but we are not sure if the potential is realized or intended.

Overall Judgment

There is definitely enough here to warrant further consideration. Communication aspects are considered, the procedures are clear, and the potential for student involvement is great. We’d like to see more justification for the procedures used, more attention given to sampling, and more of an idea how this fits into a whole program for assessing reading. We’d give it a “2/3” on a scale of 1–5, where 1 is weak and 5 is strong.

Source: All critiques except that for Sample 1 adapted from *Practice with Student-Involved Classroom Assessment* (pp. 378-382, 386-395), by J. A. Arter & K. U. Busick, 2001, Portland, OR: Assessment Training Institute. Copyright © 2006, 2001 by Educational Testing Service. Adapted by permission.