

The *Authentic Assessment Toolbox*: Enhancing Student Learning through Online Faculty Development

Jon Mueller
Professor of Psychology

North Central College
Naperville, IL 60540

jfmuller@noctrl.edu

Abstract

*To support learning about assessment for all educators, I wrote, created and published online the *Authentic Assessment Toolbox*, a how-to text on creating authentic tasks, rubrics and standards for measuring and improving student learning. The site can assist faculty development by exposing educators to the process and rationale for an alternative (authentic) model of assessment in which students are asked to perform real-world tasks which demonstrate meaningful application of essential knowledge and skills, and by describing and providing examples of how to construct such authentic assessments. Additionally, the *Toolbox* can benefit student learning by suggesting methods for promoting student engagement in substantial learning that connects to real-world applications students will recognize and value, and by describing tools (e.g., rubrics) that students can apply to their own work to gauge progress and achievement.*

Introduction

It is an interesting paradox that higher education, through the theory, research and writing of its faculty, often provides many of the frameworks for innovation in K-12 education, yet often trails behind elementary and secondary education in adopting many of these innovations. Assessment is a good example. Although educational researchers and theorists in higher education promoted alternatives to the traditional approach to assessment still commonly found throughout our entire educational system, the systematic development of standards (or outcomes, as they are commonly referred to in higher education) and alternative assessments aligned to those standards is only recently being considered on a national level in higher education, and partly only because of significant prodding from external agencies (e.g., accrediting bodies).

The development of standards has been accompanied by an increased interest in reconsidering the types of assessments that will measure those statements of what students should know and be able to do. The use of more traditional assessments (e.g., multiple-choice tests) reflects a philosophy that education is about students acquiring a certain body of knowledge and skills necessary to become productive citizens. Alternatively, authentic assessment (or performance assessment) is predicated on the belief that students need to learn how to perform the meaningful tasks they will encounter as citizens, workers, etc. In other words, acquiring a body of knowledge and skills is not sufficient. Authentic learning and assessment emphasizes students'

need to learn and subsequently demonstrate the ability to *apply* the knowledge and skills in real-world or authentic contexts.

As more educators in higher education, following the lead of K-12, come to value authentic standards and assessments, more resources are needed to assist faculty in the development of such tools. Thus, I wrote, created and published online the [Authentic Assessment Toolbox](#) to support learning about assessment for all educators.

What is Authentic Assessment?

Authentic assessment is a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills. Or, as Grant Wiggins (1993) describes it, authentic measures are “engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively. The tasks are either replicas of or analogous to the kinds of problems faced by adult citizens and consumers or professionals in the field.” Authentic tasks can range from analyzing a political cartoon to making observations of the natural world to computing the amount of paint needed to cover a particular room to performing in a chorale.

Similarly, authentic tasks can range from elaborate projects spanning several weeks to brief activities. I have heard numerous teachers mistakenly equate authentic assessment with extensive assignments requiring considerable investment of time and effort for teacher and student alike. Yet, adults often face many simpler and briefer tasks in their work or life for which we can prepare our students. For example, I prepare my introductory psychology students to evaluate the many claims they will encounter in the media. I want my students to be able to distinguish causal from correlational claims and to understand what types of evidence are necessary to support either type of claim. Thus, I created a web page ([click here](#) to see the page) which lists headlines taken from scientific news stories reported in the media. If students click on a headline it will take them to the story. I can use such a resource in a variety of ways to capture activities adults engage in on a regular basis. For example, first I ask students to determine if a headline (e.g., “Low self-esteem ‘shrinks brains’”) is causal or correlational in nature. Then, I ask them to determine if the research described in the article actually justifies such a claim. (Fortunately, as is often the case, the research in the self-esteem article was *not* consistent with the headline’s claim!)

All of these tasks replicate real-world challenges, and student performance on all of them can be assessed. Multiple-choice questions can be designed to capture some ability to apply or analyze concepts, but filling in the corresponding circle on a scantron sheet does not begin to have the face validity of asking students to complete engaging tasks that replicate real world ones (Wiggins, 1993). Reviews of research across a number of learning domains have discovered that students need to demonstrate application of their learning to effectively document the acquisition of valuable skills such as summarizing and generating and testing hypotheses (e.g., Marzano, Pickering, & Pollock, 2001; Pellegrino, Chudowsky, & Glaser, 2001).

Of course, capturing a more authentic performance does not insure validity. A measure cannot be valid if it does not effectively address the learning goals it was designed to assess. Thus, the development of good assessments of any type begins with the development of meaningful goals and standards. Although the definitions of these terms vary in use, learning goals are often written as rather broad statements to define what students should know and be able to do at some point in time (e.g., the end of 10th grade or the end of a course on music theory). Goals are subdivided into standards which are thus written more narrowly and typically in language that is more amenable to assessment to capture what students should know and be able to do at the end of a unit or a chapter or a course. Standards can be further delineated into objectives which are written even more narrowly to describe the outcomes students should achieve at the end of a particular lesson.

For a given standard, an educator would ask, "What task (or tasks) might I ask students to perform that would demonstrate that they have met that standard?" Next, the faculty member would ask, "What are the essential characteristics of good performance on that task?" Those characteristics become the criteria by which one would judge student performance. Finally, the educator would identify likely levels of performance along which he/she could judge student performance for those criteria. The criteria and accompanying levels of performance are then usually combined into a rubric, a scoring scale for the assessment. ([Click here](#) to see an example of an authentic assessment that includes 1) the standards being assessed, 2) a description of the authentic task, 3) a list of the criteria, and 4) a rubric to evaluate student performance on the task. Elaboration of the [four steps of developing an authentic assessment](#) I just outlined, along with more [examples](#) of authentic assessments, can be found at the [Authentic Assessment Toolbox](#) website.)

Why Do It?

Authentic assessments are direct measures

We do not just want students to *know* the content of the disciplines when they graduate. We, of course, want them to be able to *use* the acquired knowledge and skills in the real world. So, our assessments have to also tell us if students can apply what they have learned in authentic situations. If a student does well on a test of knowledge we might infer that the student could also apply that knowledge. But that is rather indirect evidence. I could more directly check for the ability to apply by asking the student to use what they have learned in some meaningful way. If I taught someone to play golf I would not check what they have learned with just a written test. I would want to see more direct, authentic evidence. I would put my student out on a golf course to play. Similarly, if we want to know if our students can interpret literature, calculate potential savings on sale items, test a hypothesis, develop a fitness plan, converse in a foreign language, or apply other knowledge and skills they have learned, then authentic assessments will provide the most direct evidence.

Authentic assessments capture the constructive nature of learning

A considerable body of research on learning has found that we cannot simply be fed knowledge. We need to construct our own meaning of the world, using information we have gathered and were taught and our own experiences with the world (see, e.g., Bransford, Brown & Cocking, 2000; Brown, Collins, & Duguid, 1989; Pellegrino, Chudowsky, & Glaser, 2001). Thus, assessments cannot just ask students to repeat back information they have received. Students must also be asked to demonstrate that they have accurately constructed meaning about what they have been taught. Furthermore, students must be given the opportunity to engage in the construction of meaning. Authentic tasks not only serve as assessments but also as vehicles for such learning.

Authentic assessments provide multiple paths to demonstration of learning

We all have different strengths and weaknesses in how we learn. Similarly, we are different in how we can best *demonstrate* what we have learned (Pellegrino, Chudowsky, & Glaser, 2001). Regarding the traditional assessment model, answering multiple-choice questions does not allow for much variability in how students demonstrate the knowledge and skills they have acquired. On the one hand, that is a strength of tests because it makes sure everyone is being compared on the same domains in the same manner which increases the consistency and comparability of the measure. On the other hand, testing favors those who are better test-takers and does not give students any choice in how they believe they can best demonstrate what they have learned.

Thus, it is recommended that multiple and varied assessments be used so that 1) a sufficient number of samples are obtained (multiple), and 2) a sufficient variety of measures are used

(varied) (Wiggins, 1998). Variety of measurement can be accomplished by assessing the students through different measures that allows you to see them apply what they have learned in different ways and from different perspectives. Typically, you will be more confident in the students' grasp of the material if they can do so. But some variety of assessment can also be accomplished *within* a single measure. Authentic tasks tend to give the students more freedom in how they will demonstrate what they have learned. By carefully identifying the criteria of good performance on the authentic task ahead of time, the teacher can still make comparable judgments of student performance even though student performance might be expressed quite differently from student to student. For example, the products students create to demonstrate authentic learning on the same task might take different forms (e.g., posters, oral presentations, videos, websites). Or, even though students might be required to produce the same authentic product, there can be room within the product for different modes of expression. For example, writing a good persuasive essay requires a common set of skills from students, but there is still room for variation in how that essay is constructed.

How can the *Toolbox* Assist Faculty Development?

I initially created the *Authentic Assessment Toolbox* with K-12 educators in mind. But the focus and content is broad enough to apply to assessment development at any level. I wanted to create a resource that would be accessible to educators with little or no background in assessment as well as more experienced practitioners. Feedback suggests that readers of all types have found it accessible, readable and informative.

One benefit the *Toolbox* provides faculty is exposure to an alternative to more traditional assessments. Educators who have relied extensively on tests to measure student achievement often feel a significant element of student performance is being missed. However, most teachers are not fluent with the variety of assessment options available to them. As noted by others (e.g., Guskey, 2003), K-12 educators typically have not received sufficient training in assessment development, and those of us teaching in higher education have received even less. Additionally, educators are often concerned with losing the objectivity in grading that multiple-choice and true-false tests provide, or that grading alternative forms of assessment is too time-consuming.

All of these are legitimate concerns. So, on the *Toolbox* home page I recommend that the "What is it?" chapter is a good place to start. Before other concerns can be addressed, educators need to be familiar with what is possible. I also provide a glossary to familiarize them with the language of authentic assessment. To address concerns regarding the subjective grading of authentic assessments, the *Toolbox* discusses and illustrates how authentic assessments can be designed to increase their reliability and validity. The more subjective judgment required by instructors in evaluating authentic work usually will require more time than grading an objective test. But, as I mention in the *Toolbox*, the choice is not one of either/or -- traditional *versus* authentic assessments. Some combination of traditional and authentic assessments may best serve assessment purposes.

A second benefit the *Toolbox* provides faculty is careful and detailed guidance on *how* to create an authentic assessment. Once an educator is sufficiently familiar with the concept of authentic assessment, the *Toolbox* provides a detailed tutorial on how to construct an authentic assessment using a four-step process. A [graphic representation](#) of the flow of those steps aids teachers in getting a clear sense of the rationale as well as the process. Then, each step is presented in detail with an ample number of examples. From many years of experience teaching a graduate course in authentic assessment as well as consulting with schools and districts, I have anticipated most of the questions and obstacles that teachers might encounter in creating such assessments.

Visitors to the *Toolbox* will find a separate section filled with examples from a variety of disciplines. Additionally, numerous examples are provided throughout the text to illustrate each

step of the process. Although we may be capable of grasping the abstract principles involved in assessment development, adults also benefit from concrete examples.

Furthermore, I extend the modeling of the development of authentic assessments through “workshops.” (Currently, only [a workshop on developing standards](#) is available at the *Toolbox*, but I expect to add more.) Although I am only a “virtual” guide through this process, I wanted to make the experience as personal and accessible as possible. The standards workshop allows readers to “look over my shoulder” as I assist another educator with the task of writing a learning standard. I have tried to capture what is most valuable when I work individually with teachers: the back and forth exchange of questions and ideas that evolve into learning and a product. The workshops are particularly designed for newcomers to authentic assessment as I begin the workshop with a rather naïve response from the educator and eventually guide her to a better and more informed product.

Apparently, many educators have recognized such benefits from the *Toolbox* as I have received numerous requests for permission to include portions of the *Toolbox* in training sessions for educators at all levels, including the Department of Education for the State of Hawaii and the New Jersey Department of Environmental Protection/Education Office. Additionally, the Small School Project, funded by the Bill and Melinda Gates Foundation, asked for permission to include two chapters of the *Toolbox* in its publication, *Performance Assessment*, which was distributed to several hundred high schools nationwide.

How can the *Toolbox* Benefit Students?

Ultimately, we want to ask how an educational program or resource will benefit student learning. From my experience, successful completion of each of the four steps I have identified in the development of authentic assessments should benefit teachers *and* students. For example, in the first step, if teachers more explicitly articulate their goals and standards they will be more likely to clearly communicate such goals to their students. Learners are more satisfied with the process and perform better when the goals are clear (e.g., Brophy, 1987).

Similarly, when teachers effectively design an engaging, authentic task that is appropriately aligned with a learning standard, students will discover multiple entry points to the material which allows and encourages them to connect prior knowledge and experience to the new material in meaningful ways. As a result, students can begin constructing their own meaning to develop deep and substantial understanding of concepts or skills (Pellegrino, Chudowsky, & Glaser, 2001).

Furthermore, an authentic task is more likely to address student concerns expressed in the common question “When are we ever going to use this?” It is not very often in life outside of school that we are asked to select from four alternatives to indicate our proficiency at something. Tests offer these contrived means of assessment to increase the number of times students can be asked to demonstrate proficiency in a short period of time. More commonly in life, as in authentic assessments, we are asked to demonstrate proficiency by doing something. Students will be able to see the direct application of their learning on an authentic task.

Well-designed traditional assessments (i.e., tests and quizzes) can effectively determine whether or not students have acquired a body of knowledge. Thus, as mentioned above, tests can serve as a nice complement to authentic assessments in a teacher's assessment portfolio. Furthermore, we *are* often asked to recall or recognize facts and ideas and propositions in life, so tests are somewhat authentic in that sense. However, the demonstration of recall and recognition on tests is typically much less revealing about what we really know and can do than when we are asked to construct a product or performance out of facts, ideas and propositions. Authentic assessments often ask students analyze, synthesize and apply what they have learned in a substantial manner, and students create new meaning in the process as well.

Regarding the third step, if teachers more explicitly define the criteria by which they evaluate student performance on an assignment and communicate those criteria at the beginning of the process, students will better understand what is expected of them and better how to complete the task. Again, when goals are clearer and when what is required to meet those goals is made explicit students will find the task more approachable and worthwhile. Well written criteria describe observable and measurable behaviors that reflect the most essential characteristics of good performance on a task, are clearly and briefly stated, and are written in language that students will understand.

Teacher expectations for an assignment will be further clarified by sharing a well developed rubric at the beginning of a task or assignment. For example, the rubric in Fig. 1 clearly lays out the criteria and levels of performance for an elementary (grades 3-5) science lab. If students receive this rubric before the task begins they will understand the teacher's expectations for a good lab report. Furthermore, with a well-designed rubric for a task, students can more effectively evaluate the progress they are making on that task and can better evaluate the quality of their product when they finish. Clear criteria and levels of performance make peer feedback and evaluation more likely and more valuable as well.

| Criteria | Limited | Acceptable | Proficient |
|-------------------------------|--|---|--|
| made good observations | observations are absent or vague | most observations are clear and detailed | all observations are clear and detailed |
| made good predictions | predictions are absent or irrelevant | most predictions are reasonable | all predictions are reasonable |
| appropriate conclusion | conclusion is absent or inconsistent with observations | conclusion is consistent with most observations | conclusion is consistent with observations |

Fig. 1. Rubric for an elementary (grades 3-5) science assignment.

What are the Benefits of Publishing the *Toolbox* Online?

Some of the benefits of the *Toolbox* described above could have been accomplished through a more traditional print version of this resource. Yet, I believe the online publication of the text goes beyond what a print version could have contributed in several ways. First, I was surprised by the number of readers I have been able to reach. The *Toolbox* receives approximately 15,000 hits from more than 6,000 unique visitors each month. Second, the text is much easier to find than a printed book would be. A quick search in Google has led many a reader to my site. Consequently, I have received e-mail from educators all over the world, many of whom I am sure would never have seen a print version of the text. The e-mail has provided good dialogue with other professionals as well as opportunities for future collaboration. Additionally, more than 200 educational sites link to the *Toolbox*, making it a very accessible resource.

Third, by publishing the text online, I am able to revise, update and add to the site much more frequently and easily. I do not have to wait for a second edition. I could also put the text online before it was "finished," adding to it as I went, but making the first chapters available as soon as they were written. Finally, I will be able to create a more interactive resource. I intend to add

hyperlinks to other resources that complement or supplement the *Toolbox*, and I would eventually like to add interactive exercises to enhance this learning tool.

Thus, I am left with a unique challenge that print publishing does not usually face: When do I stop? Once something is published in print the work of creating the text is essentially finished. However, once something is published on the Web, viewers expect that the resource will remain current, or be enhanced, or, at the very least, remain accessible. Fortunately, in the case of the *Toolbox*, I intend to revise, expand and enhance the site. I have received a considerable amount of useful feedback that will inform the *Toolbox's* future development. More significantly, I have received ample evidence that educators all over the world have benefited from its presence on the Web, and, as a result, that student learning will be enhanced.

References

- Bransford, J.D., Brown, A.L., & Cocking, R. (Eds.). (2000). *How people learn: Brain, mind, experience and school, expanded edition*. Washington, DC: National Academies Press.
- Brophy, J. (1987). Synthesis of research on strategies for motivating students to learn. *Educational Leadership, 45*, 40-48.
- Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Research, 18*, 32-42.
- Guskey, T.R. (2003). How classroom assessments improve learning. *Educational Leadership, 60*, 6-11.
- Marzano, R.J., Pickering, D.J., & Pollock, J.E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: ASCD.
- Pellegrino, J.W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.
- Wiggins, G.P. (1993). *Assessing Student Performance*. San Francisco: Jossey-Bass Publishers. (p. 229)
- Wiggins, G.P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.