

PERFORMANCE ASSESSMENT

Showing What Students
Know and Can Do

Susan M. Brookhart



Table of Contents

| | |
|--|-----|
| About the Author | iv |
| INTRODUCTION | |
| What Is Performance Assessment? | 1 |
| CHAPTER 1 | |
| Match Performance Assessment to Learning Outcomes: Task Design | 15 |
| CHAPTER 2 | |
| Match Performance Assessment to Learning Outcomes: Rubric Design | 35 |
| CHAPTER 3 | |
| Performance Assessment of Simple Processes | 61 |
| CHAPTER 4 | |
| Performance Assessment of Simple Products. | 69 |
| CHAPTER 5 | |
| Performance Assessment of Complex Processes | 81 |
| CHAPTER 6 | |
| Performance Assessment of Complex Products | 89 |
| CHAPTER 7 | |
| What About Group Work? | 103 |
| CHAPTER 8 | |
| Making Performance Assessment a Routine Part of Your Instruction | 111 |
| RESOURCE A | |
| The Revised Bloom’s Taxonomy. | 123 |
| RESOURCE B | |
| Description of Webb’s Depth of Knowledge Levels by Subject | 129 |
| RESOURCE C | |
| Assessment Templates. | 135 |
| Glossary | 141 |
| References | 143 |

INTRODUCTION

What Is Performance Assessment?

I VIVIDLY REMEMBER A PERFORMANCE ASSESSMENT from my early secondary school days. Year 7, to be specific. I was taking a class called home economics – today it would be called family and consumer science. For a cooking unit, we worked in groups of four. We had already learned to make different recipes and we had learned other skills like how to clean up a kitchen after cooking. It was time to put it all together and the assignment was this: plan a lunch menu; invite four teachers; and cook, serve and clean up the lunch. Daunting! We had never cooked for adults before. Plus, to serve a meal, everything has to be ready at the same time and you do not get a second chance if you burn or ruin something.

But we did it. Our group served hamburgers and green beans for lunch with chocolate chip biscuits for dessert. Our teachers said they enjoyed the meal – although, thinking back, I am not sure what else they would have said. Anyway, that was a watershed for me. I never dreaded putting together a whole meal again and I have become a pretty good cook. In my mind, that was the day it all started.

Not every performance assessment will become a watershed in a student's life. But every performance assessment *does* share something with my lunch-preparing experience – namely, what is assessed is something students *do* and the assessment is based on observation and judgement of either the process of them doing it or the finished product. In the case of our lunch, both the process (how we prepared the food, served it and cleaned up) and the product (how the food itself tasted) were observed.

Many performance assessments share another feature in common with our lunch. Performance assessments allow students to show what they know in a more authentic or realistic setting than a test does. Not all performance assessments do that, so it is important to distinguish between performance assessment, meaning assessment of a process or product using observation and judgement, and authentic assessment, meaning assessment in a real-world or near-real-world context. Some people mistakenly use *performance assessment* and *authentic assessment* as synonyms and they are not.

But that said, a nice feature of performance assessment is that it does allow students to *do* something with their knowledge. It helps show students that the purpose of their learning is more than just knowing something for a mark on a test. A carefully designed performance assessment can show teachers how students can transfer what they know to situations beyond the one in which they learned it.

Why Performance Assessment and Why Read This Book?

Performance assessment as a method has two parts: (1) students create a product, demonstrate a process or both; and (2) student performance is evaluated with observation and judgement based on clear criteria. This book brings together in one place information about how to design and use performance assessment, including what performance assessment is and what learning and assessment purposes it best serves, how to design and use performance tasks that simultaneously teach and assess learning and how to design and use rubrics (and, for special purposes, other point schemes) that support both formative and summative (marked) assessment of student learning.

This book is intended mostly for practising educators. Teachers, groups of teachers who write assessments together and principals and instructional coaches who supervise teachers will all benefit from understanding more about how to design effective performance assessments. Specific benefits from using well-designed – as opposed to poorly designed – performance assessment in class include:

- Students get a clearer vision of what they are supposed to know and be able to do, because a task provides a stronger image or conception of a learning goal than a test does. A clear vision of the goal helps with both motivation and achievement.
- Students are actively engaged in producing a product or performing an act. Active engagement with the task promotes student engagement with the learning.
- Teachers and students get information about their achievement that is more closely aligned with complex standards. For example, The Australian Curriculum: English content description for Creating Texts in Year 3 identifies that students, “*Plan, draft and publish imaginative, informative and persuasive texts demonstrating increasing control over text structures, language features, images and multi modal elements appropriate to the audience and purpose*” (ACARA, 2015). Only a series of performance assessments allows students to show they can meet this standard. No amount of testing can demonstrate this competence.
- Since the rubrics used in performance assessment require the articulation of criteria and performance standards, a description of “good work” is

available – in the classroom, to guide feedback as well as marking and for parents and others, to describe what their students can do. Contrast this with tests, where the “criterion” is often a score (e.g. 80 per cent) that carries no information about what, in fact, students know or can do. The focus of assessment more easily moves from “what did you get?” to “what did you learn?”

Of course, the students are the ultimate beneficiaries, as well-designed performance assessments make clear to students what it means to use a skill or understand a concept well enough to be able to apply it, help them learn how to do that and provide evidence of their achievement.

This book focuses on classroom performance assessments. These can be used for formative or summative purposes – that is, to inform learning while it is happening or certify and report learning after it has occurred. Large-scale assessments can use performance assessment as well. For example, a state writing assessment is a large-scale performance assessment. Both classroom and large-scale assessments are part of a balanced assessment system. The focus in this book is on the classroom because that is where learning occurs and where teachers’ assessment practises can make a difference for student learning.

Definition of Performance Assessment

The definition of *performance assessment* I use in this book is as follows (see also Brookhart & Nitko, 2015; Chappuis, Stiggins, Chappuis & Arter, 2012):

A performance assessment is one that (a) requires students to create a product or demonstrate a process or both and (b) uses observation and judgment based on clearly defined criteria to evaluate the qualities of student work.

I alluded to this earlier, but it is nice to have a formal definition. Some assessments, such as my Year 7 lunch, are clearly instances of performance assessment. Some assessments, such as a multiple-choice quiz, are clearly not performance assessment. But assessments exist along a continuum of how much construction (what has been called “doing”) is needed (Bennett & Ward, 1993). Short-answer questions on tests require more student construction than multiple-choice and other selected-response items, but they are not considered performance assessment. On the other hand, essays and other written products usually *are* considered a type of performance assessment. Where does one draw the line between an essay test question and a performance assessment?

For purposes of this book, when an assessment in the form of an essay or any other student construction (e.g. a drawing) is administered in a context that is not a test, it is treated as a performance assessment. If a series of short essay questions are asked on a test for which students sit during a designated test or examination time, I

consider those test questions. If an essay is assigned as a stand-alone project with the expectation that students would use at least some parts of a writing process (maybe brainstorming or outlining first or writing a draft and then a revised final copy), I call it a performance assessment. I consider a maths problem on a test to be a test item, while a complex maths problem or problem set that requires students to show their work and explain their reasoning is a performance assessment. This is a somewhat artificial distinction and the reason is that, in fact, the amount of construction or performance questions and tasks require is a continuum (Bennett & Ward, 1993).

Purposes for Which Performance Assessment Is Best Suited

Performance assessment is best suited to situations where you want to know not just what the student *knows*, but also what the student can *do* with that knowledge. Think of my lunch. It is one thing to know what a medium-rare hamburger looks like and quite another to produce one. Think of an primary school student learning to count by 2s. It is one thing for students to know that means they skip every other number or they say the even numbers or however they understand that. It is another thing for them to use that knowledge to correctly count out loud by 2s to 100. Think of a secondary school student learning to read, understand and appreciate Shakespeare. It is one thing for them to read and comprehend *Romeo and Juliet*, perhaps answering comprehension questions in class or on a test and another thing to use that understanding to do a complex analysis resulting in a report comparing *Romeo and Juliet* to the young lovers in *West Side Story*.

Performance assessment is also well suited to physical skills. In school settings, such skills usually have a cognitive component as well. A health and physical education teacher might observe students' skills at dribbling a basketball. Dribbling exercises also help students understand how dribbling is useful for moving a basketball around the court (knowledge), but they cannot actually use that knowledge to play basketball games until they can, in fact, dribble.

A music teacher might observe students' skills at playing a C-major scale on the recorder. It is one thing for students to know what the notes in the C-major scale are and what fingering will produce them, but they also need to be able to play the scale accurately, smoothly and with good intonation. The music teacher's observation in this case would entail both listening (to see that the notes are correct and judge the intonation) and watching (to see that the students use proper fingering).

Advantages and Disadvantages of Performance Assessment

The big advantage of performance assessment is that it furnishes information about students' abilities to use their knowledge, as I just described. Performance assessments also allow you to see how students use sets of knowledge or skills

together, not just one skill at a time. For example, students may show they can do that comparison of Romeo and Juliet and Tony and Maria and also that they can state their conclusions and argue for them using support from the literary works, producing a coherent theme.

Performance assessment has some disadvantages and most of them are related to sampling issues. If you think, for example, of “everything you want students to know about *Romeo and Juliet*”, including not only the play itself but also the larger issues of the role of theatre in Shakespeare’s day, the use of poetic language in the play and its effects on our language today and the theme of tragic romance in literature, you end up with quite a hefty amount of content. An assessment should sample that content. Your assessment will be made up of some but not all of the possible questions or tasks you could ask in that content domain. Then from students’ performance on the set of questions or tasks in your assessment, you generalise how well they have mastered “everything you want students to know about *Romeo and Juliet*”. If they score “proficient” on your assessment, you conclude they are “proficient” at the learning goals associated with the *Romeo and Juliet* unit.

Think, then, about how tests and performance assessments sample that “everything” content domain differently. Tests can cover much more content because you can ask a lot of questions – maybe 30 or 40. Performance assessments cover much less content. You may only assign one task, which in our running example is the written theme or report. There are trade-offs. Performance assessment samples some content from the “everything” domain, but deeply (if the performance assessment is well designed). Tests typically sample much more content from the domain, but less deeply. The kinds of generalisations you can make about what students know and can do, then, are different. Performance assessment results allow you to make claims about how students can use what they know, often transferring knowledge to a new situation, but in a limited content domain. Test results allow you to make claims about how students know a wider and more representative set of facts and concepts, but not whether they can use it. When your assessments are matched with your learning outcomes, you almost always need to make claims about both kinds of student achievement because most curricula and the standards on which they are based require it.

Getting Comfortable With Observation and Professional Judgement as an Evaluation Method

Sometimes teachers get uncomfortable when they have to use rubrics or other judgement-based methods of appraising student work. They are *subjective* methods of scoring as opposed to *objective* methods. While that is true, it does not follow that subjective methods are bad and objective methods are good.

An objective scoring method is one where anyone using the same scoring key, without any clerical errors, would come up with exactly the same score for a student’s

work. Think of the scoring keys you have used for multiple-choice or true/false tests. You could give the key to another person who did not know anything about the content of the test and he could score the tests accurately. But think about *what* the person is scoring – right and wrong test items. As argued previously, sometimes a complex performance assessment is a better match to the learning outcomes you need to assess.

Subjective scoring requires judgement and it is true that two scorers might not agree on the quality of a piece of work. It is the only way to score complex performances where gradations of quality will show. The important thing is that the criteria are clear, well described and matched exactly to the learning outcomes assessed and that the scorer (usually the teacher) knows what those qualities look like in student work. In other words, it is *professional* judgement, informed by clear criteria and clear understandings of students and content. Judgement in this context is a *good* thing.

Some teachers, especially at the secondary school level, have told me that they try to stay away from rubrics so they can defend students' marks more easily. (A teacher might say, "Your student got 10 out of 20 right, so that's 50 per cent and that's an F.") I understand that inclination, having faced a couple of angry parents myself, but denying students the opportunity to do complex, performance-based work is not the answer. Rather, developing clear criteria and a repertoire of ways to explain them to students and parents is the way to go.

Some teachers, especially primary school teachers, make their rubrics more objective by writing poor – but in their eyes countable and thus more easily defensible – rubrics. A rubric might include different amounts of points for "three facts included", "two facts included" and so on. This is worse than not using performance assessment at all, because it turns what should be a complex performance into a series of tick boxes – essentially a bundle of test items that are mostly about following directions and not applying knowledge. I discuss this in more detail in chapter 2.

If you use performance assessment appropriately and apply rubrics professionally and if you share the rubrics with students and help them learn what the criteria are and why they describe good work, you should be able to move subjective scoring into your comfort zone and accomplish it well. That still will not make it fun to have to talk with an angry parent, but it will position you on solid ground. It will give your students the opportunity to apply what they know in meaningful and motivating assessments.

Rationale and Research Base for Performance Assessment

Performance assessment began to "pick up steam" as an assessment method in the late 1980s and 1990s, as the standards movement got under way. In the minimum-competency era of the 1970s, the assessment focus was on basic skills (Brookhart, 2013a).

Research began to show that students were taught to recall facts and concepts but not how to apply them (Lane, 2013; Lane & Stone, 2006; Lane & Tierney, 2008). Beginning in the 1980s, schools adopted more complex standards. Performance assessment was a tool in this standards-based reform movement because performance assessments are appropriate for measuring students' skills at applying knowledge and solving problems in realistic – or somewhat realistic – contexts. Since the 1990s, performance assessment has taken its place as one (tests are the other) of the mainstream methods of assessing student learning.

To date, performance assessment, rather than large-scale assessment, has proven more successful in the classroom (Tung, 2010). But even in the classroom, teachers find performance assessment challenging to design and implement (Dietel, 1994). Defining specific performance criteria and using them systematically is one of the keys to successful implementation of performance assessment (Herman, Aschbacher & Winters, 1992; Lane & Tierney, 2008).

There is some evidence that when students understand the criteria on which they will be judged, they improve their performance. This research has been conducted in a variety of subjects and year levels: primary projects (Higgins, Harris & Kuehn, 1994); primary and middle years writing (Andrade, Du & Mycek, 2010; Andrade, Du & Wang, 2008; Coe, Hanita, Nishioka & Smiley, 2011); primary mathematics (Parker & Breyfogle, 2011); middle years mathematics (Ross, Hoagaboam-Gray & Rolheiser, 2002); middle years special education (Lee & Lee, 2009); and secondary school humanities and social sciences (Panadero, Tapia & Huertas, 2012; Ross & Starling, 2008). On balance, though, the research evidence is mixed and more research needs to be done (Brookhart & Chen, 2014; Johnsson & Svingby, 2007). It is clear that setting students tasks that require them to apply what they know and providing clear rubrics *can* measure important, non-testable learning outcomes (Lane, 2013; Lane & Stone, 2006; Lane & Tierney, 2008). It is equally clear that this does not always happen. It is just as possible to produce a low-quality performance assessment as it is to produce a low-quality test. Part of the purpose of this book is to help you design performance assessments that measure *exactly* the learning outcome you need to assess and in the process, develop students' capacities to use what they know to solve problems and apply their knowledge in contexts beyond the ones in which they learned.

Different Types of Performance Assessment

Performance assessments come in a variety of types. Not surprisingly, the type of performance assessment you select should match your purpose for assessment. This section presents brief descriptions of some different types of performance assessment, with examples and an explanation of how the rest of the book is organised so you can learn to use all of them.