




1-1-1994

Measurement and judgment in curriculum assessment systems

Glen Rogers

Alverno College, grogers@pacific.edu

Follow this and additional works at: <https://scholarlycommons.pacific.edu/provost-facarticles>

 Part of the [Curriculum and Instruction Commons](#), [Educational Administration and Supervision Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Higher Education Commons](#)

Recommended Citation

Rogers, Glen, "Measurement and judgment in curriculum assessment systems" (1994). *Office of the Provost Faculty and Staff Articles*. 6. <https://scholarlycommons.pacific.edu/provost-facarticles/6>

This Article is brought to you for free and open access by the Office of the Provost at Scholarly Commons. It has been accepted for inclusion in Office of the Provost Faculty and Staff Articles by an authorized administrator of Scholarly Commons. For more information, please contact mgibney@pacific.edu.

**MEASUREMENT AND JUDGMENT IN
CURRICULUM ASSESSMENT SYSTEMS**

Glen Rogers

MEASUREMENT AND JUDGMENT IN CURRICULUM ASSESSMENT SYSTEMS

Glen Rogers

While educators show increasing interest in performance assessment, many measurement specialists express alarm over implications for increased costs. When we see ourselves as assessment practitioners who are collaboratively working with educators to create integrated assessment systems, however, we must first rethink relationships among our measurement, curriculum development, and institutional assessment goals before we can understand what assessment costs are implied. The quality of the contribution of assessment practitioners depends on their knowing when to pursue measurement, and when instead to facilitate coordinated human judgment.

I mean this conclusion to guide rather than exclude contributions to assessment by educational researchers, evaluators, or measurement specialists. I have worked for ten years as an assessment practitioner both within Alverno College and in collaboration with others pursuing assessment at diverse institutions, and I have needed to cycle through each of these research, evaluation, and measurement roles to address each assessment system and its coherence.

I have found that higher education faculty and administrators want an assessment system that helps them create coherence among all aspects of a diverse curriculum and that unites learning and assessment. Faculty want to teach and assess abilities that make a difference for students well beyond college; they want to be able to explain these abilities to students and demonstrate them to employers and the public at large. Performance-based assessment can synergistically advance these diverse educational reform goals when faculty carry out a comprehensive and coherent performance assessment system that ensures student learning of complex, multidimensional abilities. Assessment practitioners significantly support this work by connecting assessment at the level of the student with institutional level assessment.

In curriculum-integrated assessment systems, institutional- and student-level assessment *each* accommodates the other so that the system: (1) supports students to achieve further learning, (2) supports curriculum improvement, (3) reflects depth and breadth in institutional outcomes, and (4) contributes to informed judgment about the value and worth of institutional programs.

When we accept the challenge of integrating student and institutional assessment across the curriculum, we must rethink and broaden the meaning of validity. First, in performance-based curricula, student performance is often supported by feedback, which confounds measurement. Nonetheless, students and faculty learn much more from performance assessments than just the ability that is the "measurement" focus of individual assessments. For example, students learn what it means to perform, use feedback, and transfer performance across contexts; faculty learn how their teaching is connected to student learning. In contrast, measurement assumes that student abilities can be assessed and reported unidimensionally.

Second, in rethinking validity, assessment practitioners must recognize that performance is often inherently unique. Even in the same broad context, students create different ways to perform. Even when well-specified, "similar" complex tasks pull different multidimensional mixtures of abilities. Even within a single course, complex performance tasks challenge measurement of unitary constructs. Still, with multidimensional performances, assessors can give feedback that is meaningful to students because it is relevant to future performance on the job, in the community, and in further educational pursuits.

Third, when we build a community of judgment, judgment can approximate measurement ideals. Interdisciplinary faculty in an integrated curriculum develop and use coherent plans to coordinate their assessments, even though they do not generally expect to comparably *measure* the "same" ability across performance assessments. For example, Alverno's curriculum specifies levels of sophistication of broadly described abilities that students must demonstrate across the curriculum. The corporate faculty have designed curriculum-wide ability criteria to accommodate valid but different faculty perspectives on the precise meaning of an ability, as they assess students in the context of their own courses or disciplines. Common, *broad* criteria help faculty coordinate their context-embedded judgments and help students construct their own understanding of how to transfer their performance to new settings. They are not designed to yield comparable *measurements*.

A curriculum-integrated assessment system *approximates* "measurement" ideals through multiple, cumulative, time-extended observations of the student's performance across diverse performance settings. This judgment system, however, destroys the expectation that a specific ability is *measured* and exists at any one point in time. From a learning point of view, extended practice on diverse tasks helps students construct and internalize better performance strategies. So, assessment practitioners sometimes need to resist making measurement goals superordinate.

For example, Alverno students complete over 100 assessments around eight broad abilities, demonstrated within varied and specific contexts. Each context presents new challenges. Curriculum coherence provides students with a frame for understanding how to create new multidimensional performances. Similarly, faculty translate criteria and add specificity to judge performance in new contexts. For multidimensional and complex performances, we sometimes need to celebrate differences in how performances are interpreted, because acknowledging new insight counterbalances normative consensus in a vital community of judgment and validates students' self-assessment of their creative achievements.

We need to be more explicit about when to pursue measurement, when to facilitate human judgment, and how these pursuits are connected. We still need to be clearer about what "assessment" means. We should avoid equating it with measurement even though measurement validity includes measurement consequences. Fortunately, for both measurement specialists and faculty, the term "assessment" can highlight the role of informed judgment. Agreeing to be mindful of this meaning may improve communication among measurement, public policy, and higher education communities, since judgment plays such a significant role in developing and crediting student and institutional performance.

Rethinking the role of judgment and measurement affects our practice in integrated assessment systems. First, professionally trained evaluators can facilitate collaborative *judgment* processes in the design, implementation, and evaluation of assessments. For example, when assessments are intended to integrate diverse curricular experiences, designs need to be made sensitive to this diversity. As an assessment practitioner, I have gathered samples of performance and set up processes that help interdisciplinary faculty collectively benchmark their own criteria against performance. I have found that faculty can remain sensitive to various manifestations of performance in context, avoid creating a "sea of criteria," and develop a basis for similarity in judgment and understanding. Pushing for measurement properties often did not make sense in this integrated performance assessment context.

Second, educational researchers can use existing measures of theoretical constructs in the *study* of

educational outcomes and interventions, and can illuminate educational principles within the curriculum

to support curriculum development. The ideal of construct validity that is connected to a line of

"scientific" interpretative inferences does provide *one* important source of external measures and

theoretical frameworks for *studying* student learning, growth and abilities. While educational outcomes

transcend these unidimensional measures and nascent theories, there can be important points of contact

Measurement and Judgment in Curriculum Assessment Systems

page 10

with curricula when a range of measures and strategies are used.

Finally, researchers can work with faculty who are conceptualizing curriculum abilities. For

example, in collaboration with faculty and research staff, I am leading a study of alumna performance of

multidimensional abilities that identifies what curriculum abilities are still used by five-year alumnae.

Alverno researchers conducted and coded interviews designed to study performance of abilities across

naturally occurring settings, roles, and positions and in a specific context. Our findings combine the

proportion of alumnae that demonstrate abilities with (a) independent faculty judgments of effectiveness of

performance, (b) benchmarked, narrative examples of performance that illustrate frequently coded

abilities, and (c) comparative findings from similar studies of outstanding professionals who are *not*

alumnae of our college. With these results, we have been able to help demonstrate accountability to

various audiences and to also provide faculty an opportunity to explore curriculum development in

relation to their goals for current students. Our college has been able to devote resources to indepth

educational research because our faculty take responsibility for developing abilities within a coherent

curriculum.

Performance-based assessments are more time consuming relative to *some* of their individual

measurement properties. Nonetheless, educators seem to find compensating cost advantages when they

integrate them into a curriculum assessment system. Educators who fully integrate teaching, learning, and

assessment say they recover assessment costs from restructuring teaching around performance goals.

Assessments are more integral to learning activities, because curriculum goals are advanced rather than

subverted by the assessment process, and because measurement at the level of individual assessments is

less important when you have curriculum-integrated, student assessment systems. For institutional level

assessment, the cost-saving challenge is to develop strategies for aggregating performance from ongoing

curriculum-integrated assessments (e.g., sampling and systematically judging portfolios in relation to

benchmark examples that illustrate criteria).

Of course, there are working examples of performance assessments that can even support

relatively high-stakes contexts. In medicine, for example, standardized patient assessments of the ability

to make a good medical diagnosis achieve high reliability when enough performance tasks are used, and

they have been integrated into medical curricula. These commendable achievements involve substantial

and prolonged commitment of resources, but do support processes for the aggregation of data that are

accessible and credible to external accrediting bodies. At the same time, it is important to realize that

these measurement achievements are limited to relatively circumscribed professional settings. My

argument is that we should build curriculum assessment systems that teach and judge multidimensional

abilities, if we are also to pursue costly *measurement* objectives for performance-based assessments.

Author

Glen Rogers is senior research associate, Office of Research and Evaluation, Alverno College, Milwaukee,

Measurement and Judgment in Curriculum Assessment Systems

page 45

Wisconsin.