The Challenge of Knowing What Students Know

James W. Pellegrino

Department of Psychology
University of Illinois at Chicago, USA

Educators assess students to learn about what they know and can do, but assessments do not offer a direct pipeline into a student's mind. Assessing student knowledge and educational outcomes is not as straightforward as measuring height or weight; the attributes to be measured are mental representations and processes that are not outwardly visible. Thus, an assessment is a tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know. Deciding what to assess and how to do so is not as simple as it might appear. Existing guidelines for assessment design emphasize that the process should begin with a statement of the purpose for the assessment and a definition of the content domain to be measured (AERA et al., 1999; Millman & Greene, 1993). The *evidence centered design* (ECD) model at the core of the system described by Mislevy, Steinberg and Almond (this volume) goes well beyond such prescriptions. Their system attempts to provide a comprehensive picture of the structure and process of educational assessment. They simultaneously offer a challenge to the field regarding how business should henceforth be done.

Mislevy et al. are to be applauded for the completeness with which they have sketched out multiple components of a systematic approach to educational assessment. The cost, however, is that their proposed system is exceedingly complex. Understanding all the elements and entailments of their system is a conceptual challenge even though they have been careful to provide glimpses of what each component represents, its operation, and their collective interaction. Mislevy et al. also provide examples from the domain of language assessment that help concretize how things might unfold in specific cases. These cases illustrate the very important role of context and purpose in shaping the ultimate process and product of educational assessment.

Are all the elements in the Mislevy et al. system needed to make progress conceptually and operationally? Perhaps not. Many of the key ideas in the ECD model have been described in a simpler and arguably more straightforward form in the National Research Council report *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino, Chudowsky, & Glaser, 2001). In the latter, the process of

reasoning about students' knowledge from evidence obtained in an educational assessment is portrayed as a triad of three interconnected elements – the *assessment triangle*. The vertices of this triangle represent the three key elements underlying any educational assessment: a model of student *cognition* and *learning* in an academic domain; a set of beliefs about the kinds of *observations* that will provide evidence of students' competencies; and an *interpretation* process for making sense of the evidence. These three elements may be explicit or implicit, but an assessment cannot be designed and implemented without consideration of each. A major tenet of the ECD approach is that for an assessment to be effective, the multiple elements must be in synchrony. The simpler assessment triangle framework, shorn of much of the complexity in Mislevy et al., provides a useful starting place for analyzing the underpinnings of current assessments to determine how well they accomplish the goals intended, as well as for designing future assessments.

Starting with the simpler assessment triangle schema, we can readily discern that even such a simple framework poses some major challenges for assessment design and implementation. The challenges are only exacerbated in the Mislevy et al. system. Consider for example what is implied by the knowledge underlying the cognition vertex. The latter serves as the starting point for designing any assessment. Empirically based models and descriptions of the nature of knowledge and competence are lacking in many areas of the curriculum even though our knowledge of cognition has grown considerably over the past few decades (see e.g., Bransford, Brown & Cocking, 1999; Donovan, Bransford & Pellegrino, 1999; Pellegrino, Chudowsky, & Glaser, 2001). This problem gets compounded when we consider the knowledge base about the affordances and properties of tasks vis-à-vis tapping critical aspects of cognition, or the mapping of measurement and interpretive models against task features and cognitive models. Despite the many examples provided in Pellegrino et al. (2001) and those provided by Mislevy et al. for the language area, reality indicates that relatively few cases exist in which anything approaching the full analysis, design, and implementation scheme described by Mislevy et al. has been approximated.

If much of the knowledge infrastructure from which to generate all the necessary models and methods called for in the Mislevy et al. "idealized" scheme is lacking then what purpose is served by such a complex model? Does it help or hinder progress? On what grounds can we argue that the model has any basis in reality irrespective of its highly rational structure? Just what is necessary and sufficient in the Mislevy et al. scheme let alone possible? Thus, a fundamental question about the Mislevy et al. system is whether the decoupling of their multiple components is conceptually justified as well as technically feasible. At the heart of their system is the conjecture that all the elements can be isolated and properly specified. This proposition remains to

be proven by accumulating a sufficient set of actual cases. We simply don't know all the logical and psychological interdependencies that may exist among the multiple elements of the system that Mislevy et al. have initially described. Therefore, serious questions exist about the validity and utility of starting out with such an intricate and complex system.

Criticisms of the details of their system notwithstanding, we must take heed of the larger practice and policy implications that flow from an ECD model of educational assessment. Whether we like it or not, Mislevy et al. alert us to the fact that the generation of more valid educational assessments will require two things. The first is an expanded knowledge base about domains of academic knowledge and performance, about tasks and their affordances, and about measurement models that appropriately match different forms of data. The second requirement is multidisciplinary teams that work together rather than separately to bring to bear models of domain knowledge, tasks, and analysis methods in the process of designing and developing educational assessments that meet the requirements of conceptual systems consistent with the core propositions of an ECD approach.

Another issue that Mislevy et al. bring into perspective is the need for a much deeper appreciation of what is common and what is unique across educational assessment scenarios. What is common is that all assessments represent the same fundamental process of reasoning from evidence and thus all should emanate from a rich, principled, and empirically-based understanding of what it means to "know" something, including how that knowledge develops over time. Furthermore, the two other major components of assessment – the observations and the interpretation process -- should be tightly coupled to that underlying model of competence. The process of collecting evidence to support inferences about what students know represents a highly principled chain of reasoning from evidence about student learning that characterizes all assessments, from classroom quizzes and standardized achievement tests, to computerized tutoring programs, to the conversation a student has with her teacher as they work through an experiment. Failure to consider the need for synchrony among the three major elements of the assessment triangle can produce an incoherent and non-efficacious assessment. Impacting the nature of that synchrony is what is unique across assessments -- the context or purpose which influences the types of inferences that an educational assessment is designed to support thereby impacting the type of information to be obtained and how it is to be used.

Failure to appreciate both the common and unique elements of educational assessment situations can lead to fruitless debates about the inadequacies of educational assessments. Examples include the argument frequently heard from teachers about the lack of utility of standardized tests for informing instructional practice, as well as the arguments by policy makers

that tests embedded in instructional practice lack the breadth and generalizability needed to make judgments about the efficacy of larger educational programs or systems. Both arguments are typically correct, and a solution to the concerns of both groups lies not in trying to use one mode of assessment to fit multiple purposes. The system described by Mislevy et al. points out the need to match the design process to the level and types of inferences appropriate to the context and purpose of an educational assessment. Good educational assessment that meets the needs of the policymaker may not be good educational assessment to meet the needs of the teacher and vice versa.

A final issue follows from the discussion above and concerns the role of educational assessment in educational policy making. The process of testing students has taken on new meaning. Tests have become a major policy instrument in high stakes accountability environments. Despite many admonitions to the contrary (e.g., Heubert & Hauser, 1999; AERA et al., 1999), high stakes decisions about what students know – *how much they know and how well they know it* – continue to be based on the results of single tests. These decisions impact the lives of students, teachers and administrators. The Mislevy et al. paper supports the principle that no educational assessment, no matter how well designed, can or should serve such a high stakes policy function. Rather, to truly support informed decision making in areas of educational policy and practice, it is far more sensible to think in terms of systems of educational assessments that operate together to provide the levels and types of information typically desired about the nature of student knowledge and competence. Such systems should exhibit three critical properties: comprehensiveness, coherence and continuity. These three characteristics describe an assessment system that is aligned along three dimensions: vertically, across levels of the education system; horizontally, across assessment, curriculum and instruction; and temporally, across the course of a student's studies (Pellegrino, et al., 2001).

In summary, Mislevy et al. provide us with a comprehensive model for educational assessment that clearly outstrips the current knowledge base. Furthermore, it may also contain a level of specificity that lacks conceptual and empirical justification. Much more evidence will be needed about the necessity, sufficiency and utility of what they propose. Even so, the broader rationale for their system contains many important implications for assessment design, practice and policy. We have a considerable distance to travel in meeting the scientific, practical and political challenges implied by evidence centered design systems. Nonetheless, even small steps along the path of such systems will undoubtedly lead to significantly improved educational assessment.

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice.* Washington, DC: National Academy Press.

Heubert, J. P., & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion and graduation.* Washington, DC: National Academy Press.

Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement (3rd Edition)* (pp. 335-366). Phoenix, AZ: Oryx Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.