MEASUREMENT, *4*(4), 240–269

COMMENTARIES

Implications of Assessments for Learners

Theo L. Dawson
Developmental Testing Service
Northampton, Massachusetts

Kurt W. Fischer
Department of XXXX
Harvard Graduate School of Education

Patz (this issue) presents a straightforward and convincing account of the psychometric and practical considerations that should concern the designers of the No Child Left Behind (NCLB) Act of 2001 science assessments. Among these, he includes content standards. As cognitive developmental psychologists with an interest in science learning, we have a natural interest in content standards and related assessments, which ideally should be informed by scientific knowledge about the pathways and processes through which students' knowledge develops. Both standards and tests should be informed by scientific research on how students learn and develop knowledge.

Patz (this issue) makes several references to the relation between student learning and testing, making a good beginning toward an informed developmental science basis for NCLB standards and testing. He stresses the importance of adopting "sound, developmentally appropriate, realistic, and achievable content standards" (p. 201), and he discusses aligning instruction, competencies, and assessment, asserting that "assessments must be informed by all that is known about how students learn science" (p. 234).

These concerns are fundamental, but we suggest that they need to be linked more explicitly: To ensure that content standards and associated assessments are develop-

Correspondence should be addressed to Theo L. Dawson, Developmental Testing Service, 35 South Park Terrace, Northampton, MA 01060. E-mail: theo@devtestservice.com

mentally appropriate, realistic, and achievable, they must be informed by knowledge about how students actually learn particular science concepts and skills.

Although Patz (this issue) makes a number of sound suggestions for enacting these fundamentals, he neglects one essential ingredient—research into the pathways through which science skills and concepts develop. Scaling studies can yield useful information about the order in which students provide correct answers to test questions, but they cannot explain the ordering and the processes producing it. Task analyses can yield detailed accounts of the component skills required to solve a problem, but their usefulness requires that they be grounded in a research framework that specifies valid methods for task analysis related to learning. Error patterns can yield information about students' misconceptions, but research and related theory must illuminate the reasons for the misconceptions. In order to understand (a) why items order the way they do, (b) how skills and concepts are acquired, and (c) how misconceptions are constructed, educators and researchers need to systematically study the development of students' science skills and concepts (Fischer & Bidell, 2006). This research is necessary for students of all ages and grades, from preschool through college and graduate school to adults pursuing lifelong education.

Several research programs have established strong models for this kind of effort. Case, Griffin, and their colleagues, for example, have shown how cognitive developmental research can provide grounding for strong, effective mathematics curriculum and assessment (Case & Okamoto, 1996; Griffin, 2002). Dawson-Tunik (in press) devised a framework for designing and coordinating curriculum and assessment grounded in a universal ruler for cognitive development and learning.

Why is it important to bring understanding of learning pathways to test design? First, this kind of knowledge, by providing an account of patterns of knowledge construction, could help test designers to decide what to assess. Second, it could lead to improvements in item design, for example, by suggesting developmentally informed distractors for multiple choice items. Third, it could provide a solid empirical basis for setting appropriate proficiency standards. Most important, knowledge of this kind can make test results much more useful to teachers and students. If educators are to use testing to evaluate teacher effectiveness or to determine whether students are performing at an adequate standard, we owe it to students and teachers to make testing as informative as possible. The results of a good test provide teachers with specific knowledge about the next thing students need to learn in order to become proficient. This should be the central objective of testing in education.

REFERENCES

Case, R., & Okamoto, Y. (with Griffin, S., McKeough, A., Bleiker, C., Henderson, B., et al.). (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, *61*(5–6, Serial No. 246).

Dawson-Tunik, T. L. (in press). Cycles of research and application in science education. In K. W. Fischer & T. Karz (Eds.), *Building usable knowledge in mind, brain, and education*. Cambridge, MA: Cambridge University Press.
Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In W. Damon (Series Ed.) & W. Damon & R. M. Lerner (Vol. Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 313–399). New York: Wiley.
Griffin, S. (2002). The development of math competence in the preschool and early school years: Cognitive foundations and instructional strategies. In J. M. Royer (Ed.), *Mathematical cognition: Current perspectives on cognition, learning, and instruction* (pp. 1–32). Greenwich, CT: Information Age.
No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

NCLB Science Assessments: A Unique Opportunity

Jennifer L. Dunn, Brian Gong, and Scott Marion
Center for Assessment
Dover, New Hampshire

Richard Patz's (this issue) lead article provides a useful summary of several issues in test design, most of which have been discussed and implemented in several state programs. He discusses eight key components of large-scale test development (standards, blueprints, items, scoring, scaling, equating, standard setting, and reporting), outlining a number of operational options for each stage. He particularly attends to matrix sampling's benefits for addressing content standards more broadly than could be done otherwise and deals with concomitant issues such as linking nonparallel forms through clever approaches such as a fixed anchor design. Although several of these design issues and operational procedures have been discussed in contractor proposals and assessment technical reports (e.g., Kentucky and Maryland's state assessment program documentation from the early 1990s deals extensively with matrix sampling and associated issues), Patz's article provides a valuable service in introducing this information to a wider audience in a more accessible, highly readable form. Although Patz alerts readers to some important trade-offs and still-thorny challenges, a deeper treatment of these issues will still require an interested person to go to these other sources.

Correspondence should be addressed to Scott Marion, Center for Assessment, P.O. Box 351, Dover, NH 03821. E-mail: smarion@nciea.org