A Continuous Measure of Growth Outcomes: Will It Do What It Proposes to Do?

Rebecca R. Fewell

Journal of Early Intervention 2001; 24; 185
DOI: 10.1177/10538151010240030202

J. (1997). Second-generation re-
ne field of early intervention. In M.
k (Ed.). *The effectiveness of early*
(pp. 3–22). Baltimore: Paul H.

P., Warfield, M. E., Shonkoff, J. P.,
rden Krauss, M. (2001). Children
lities: A longitudinal study of child
nt and parent well-being. *Mono-*
he Society for Research in Child De-
66, (No. 3, Serial No.256).

J., & Murray, C. (1994). *The bell*
lligence and class structure in Amer-
ew York: The Free Press.

R. (2000). Assessment in early in-
and early childhood special educa-
ing on the past to project into our
ics in Early Childhood Special Ed-
43–48.

996). Assessment and its importance
ntervention/early childhood special
In M. McLean, D. B. Bailey, & M.
Eds.), *Assessing infants and pre-*
with special needs (2nd Edition) (pp.
lewood Cliffs, NJ: Merrill.

ation Goals Panel. (1999). *The Na-*
cation Goals report: Building a na-
rners, 1999. Washington, DC: U. S.
nt Printing Office.

(1945). Rejoinders and second
Symposium on operationism. *The*
cal Review 52, 291–294.

995). Some concerns about the pro-
al of Early Intervention, 19, 21–23.

spondence to **William H. Brown,**
nent of Educational Psychology, Col-
tion, University of South Carolina,
outh Carolina 29208. *E-mail:*
WM.SC.EDU

Journal of Early Intervention, 2001
Vol. 24, No. 3, 185–187

A Continuous Measure of Growth Outcomes: Will It Do What It Proposes to Do?

REBECCA R. FEWELL
University of Miami

Priest et al. (this issue) present two investigations from a national collaboration of special education early childhood researchers engaged in a multi-site, multi-year effort to develop and validate general growth outcomes for children. The first investigation identified a set of outcomes that describe the development of children birth to age 8; the second study validated the outcomes. The identification of critical growth indicators, valid over an 8-year developmental period, could be exceedingly useful. If the outcomes are to exemplify new directions in assessment and be adopted as measures of national goals, then the identification methods must be able to withstand rigorous scrutiny. The research described in Priest et al. is both interesting and puzzling. Of concern is whether the research is sufficiently meritorious to result in a valid measure of growth outcomes suitable for national use. This review raises concerns for further consideration.

The first study describes the process used to select the outcomes. The researchers divided themselves into age-related teams to approach this task. Following a review of skill inventories and curricula programs the teams selected important developmental milestones for the various ages. Subsequently, they came together and used a consensus reduction technique to narrow the list to a single set of 15 general growth outcomes that could be effective measures across the three age groups. An immediate concern is the process used to identify the outcomes. The researchers are from a single collaborative project and their discipline-related expertise is similar. Although this might lead toward amicable agreement, the downside is that it limits their perspectives on effective methods for determining growth outcomes and in interpreting the clinical significance of specific skills.

A second and related concern is the criteria the group used in narrowing the list of skills. By requiring that the critical markers be applicable across all age groups, some important skills at each age might have been excluded. This compromises the content of the outcomes. Had the researchers relied on statistical procedures to identify the critical skills across ages (if they insist on this criteria), the outcome list might be quite different and more defensible. With the excellent statistical methods available to researchers today, it is difficult to understand and justify the use of a limited set of clinical opinions as a valid method for the selection of critical skills.

A close examination of the specific outcomes raises additional concerns. The final outcomes are very familiar, not only to early intervention practitioners, but to anyone who picks up a flier on screening children in a pediatrician's waiting room or scans a parenting magazine. Given the field's familiarity with these outcomes, one is puzzled as to what is new in these outcomes. The outcomes are essentially the same skills others have been measuring for decades. Yet, the authors suggest there is no valid system presently in place to measure child growth in the five areas of development. Is it appropriate or innovative for the researchers to return to the very sources, whose adequacy they had earlier questioned, in order to identify the growth out-

comes? Further, as originally worded, some of the outcomes are likely to have been validated in national norming studies. By rewording these statements and placing them in the selected format, validity, from a measurement perspective, is lost. One must now question whether the outcomes will be sufficient to measure children's competence in a manner that will satisfy the public's outrage over the lack of accountability in the field of education.

The researchers indicate a future step is to operationalize the outcome statements. This term implies specificity and perhaps these future statements will address the measurement concerns. A logical place to search for operational descriptions of the outcomes will be the very same sources previously reviewed to identify the outcomes. It is expected that each of the 15 statements will have several subindicators. This exercise begins to look circular. The content is unlikely to change; redundancy is expected to continue. Indicators will again be borrowed from their original source and will be rewritten or adapted to the specifications of the researchers. One might reasonably question whether these outcomes, or the forthcoming operationalized statements, are the markers our nation should use for determining the preparedness of children to enter school ready to learn.

The researchers suggest that the outcomes or operationalized statements will permit them to identify instructional targets for individual children. The birth to 8-year period of development is the most rapid time of growth. Will the statements, even within a 3-year age span, be valid for children whose growth needs are likely to be vastly variant? Will these statements be sufficiently generalizable that facilitators can identify subsequent and appropriate targets for instruction even when it means moving to the next age range? Concrete, measurable behavior, such as that required in the goal statements on Individualized Family Service Plans or Individualized Education Programs, is presently missing from these outcomes. Hopefully, the future indicator statements will include accurate, specific, and measurable language so that the progress of individuals and of groups of children can be assessed through valid and reliable means.

Turning to the second investigation, the researchers describe a mail survey used to validate the general growth outcomes. To determine whether something is valid for a nation, one must have a valid national sample. From the beginning, the investigators buy in to bias: They recruit names from membership lists of national organizations, which, they acknowledge, cannot form a representative sample. The use of a mail survey to validate the outcomes is open to question. According to Kerlinger and Lee (2000), "the mail questionnaire has serious drawbacks unless it is used in conjunction with other techniques" (p. 603). Continuing their critique of questionnaires, these authors indicate that the defects in this method, a possible lack of response and the inability to verify the responses, ". . . are serious enough to make the mail questionnaire worse than useless . . . ." Kerlinger and Lee provide some concrete targets for use in measuring return rates: "If mail questionnaires are used, every effort should be made to obtain returns of better than 80–90%, and lacking such returns, to learn something of the characteristics of the nonrespondents" (p. 603). A response of less than a third of the surveys mailed to the parent group, and the decision to include the pilot sample even after altering the measure, call into question the decision to pursue a path plagued with problems.

Finally, the statistical methods used to determine the validity of the outcomes must be reviewed. A 3-point scale (*critically important, very important,* or *somewhat important*) was used by the parents and professionals to rate the statements. The pilot study should have identified this as a problem area. The researchers failed to get the desired discrimination on this likert-type rating continuum. With a scale of five or seven ratings there is a greater chance for variance than with a scale of two or three ratings. The limited response variance confounds the variance and is a threat to validity.

Is this where research is in terms of answering questions concerning validity of outcomes? Given the importance of this topic,

researchers are remiss if they do not use methodologies that have scientific rigor, such as criterion-predication validity, construct-identification validity, and confirmatory factor analysis techniques (Anastasi & Urbina, 1997). Zigler and Styfco (2000), commenting on the National Education Goal 1 and school readiness, state emphatically: "The construct of school readiness shares a fault with that of social competence: Neither has an authoritative definition and therefore both lack standard measurement .... A program must not only have a clear, attainable goal, but there must be a way to access whether it is being realized" (p. 69). Changes in children's IQ or DQ do not have to be, and hopefully will not be, the preferred measurement continuum for the attainment of growth outcomes, but the forms educational researchers do adopt, must be valid and defensible. According to Webster's dictionary (Guralnik, 1980) the word "valid" is from the Latin word, "validus" which means strong and powerful. Webster defines this word as "... having legal force; properly executed and binding under the law ... well-grounded on principles or evidence; able to withstand criticism or objection, as an argument; sound" (p. 1568). That is precisely what we need in order to provide convincing evidence of facts for our nation and our educational system. Perhaps this investigation can serve best as a pilot step in the rigorous, highly skilled endeavor that must be undertaken if our nation is to have conclusions that can be trusted.

REFERENCES

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Kerlinger, F.N., & Lee, H.B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College.

Gurnalnik, D. B. (1980). *Webster's New World Dictionary of the American Language,* (2nd ed.). New York: Simon & Schuster.

Zigler, E., & Styfco, S.J. (2000). Pioneering steps (and fumbles) in developing a federal preschool intervention. *Topics in Early Childhood Special Education,* 20, 67–70, 78.

Address Correspondence to **Rebecca Fewell,** University of Miami, School of Medicine, P.O. Box 014621, Miami, Florida 33101. *E-mail: rfewell@peds.med.miami.edu*