ELSEVIER

Early Childhood Research Quarterly 18 (2003) 547–566

Early Childhood Research Quarterly

Predicting children's academic achievement from early assessment scores: a validity generalization study☆

Juhu Kim[a,b,*,1], Hoi K. Suen[c]

[a] *Division of School Evaluation, Korean Educational Development Institute, 92-6-Umyon-dong, Socho-gu, Seoul 137-791, South Korea*
[b] *University of Missouri-Kansas City, 5100 Rockhill Road, Kansas City, MO, USA*
[c] *103 CEDAR Building, Pennsylvania State University, University Park, PA 16802-3108, USA*

Abstract

Although there have been numerous studies investigating the predictive validity of early assessment, observed predictive validity coefficients across studies are not stable. A validity generalization study was conducted in order to answer the question of whether the relationship between early assessment of children and later achievement is generalizable or situation-specific. This study examined 716 predictive correlation coefficients from 44 studies using Hierarchical Linear Modeling (HLM). The findings of this study revealed that predictive validity of early assessment is not generalizable. Additional analyses indicated that predictive validity differ across assessments as a function of test type, specific construct being assessed, length of prediction, and administration procedures. The most impressive finding in this study was the variability of effect sizes across different test administration types. In particular, tests that were scored through ratings were found to be most effective. These findings suggest that instead of addressing a broad predictive validity between a test and a criterion measure, it is necessary to understand early assessment procedures as a whole system by including considerations of various variables related to testing conditions.

Keywords: Early assessment; Predictive validity

The fundamental purpose of early assessment is to find and help children who have potential learning problems and who may need special programs (Barnett, Macmann, & Carey, 1992;

☆ An earlier version of this paper was presented at the Annual Meeting of the American Educational Research Association in New Orleans, April 2000.

* Corresponding author. Tel.: +82-2-3460-0364; fax: +82-2-3460-0152.
E-mail addresses: Kimju@kedi.re.kr (J. Kim), Hoisuen@psu.edu (H.K. Suen).
[1] Tel.: +1-816-333-9731; fax: +1-816-235-5270.

0885-2006/$ – see front matter © 2003 Published by Elsevier Inc.
doi:10.1016/j.ecresq.2003.09.011

Lidz, 1983; Tramontana, Hooper, & Selzer, 1988). Numerous screening inventories, scales, and other types of formal and informal measurement tools have been developed for this purpose. These tools include many standardized tests such as the Metropolitan Readiness Test, the Gesell School Readiness Test, DIAL-R, McCarthy Screening Test, Early Screening Inventory, and so forth; and informal approaches based on portfolio assessment and observations (e.g., Meisels, 1993; Pellegrini, 1992). Through exploring young children's developmental and learning problems, early assessment has become one of the most important tools to guide the provision of better intervention and educational services in early childhood education.

One of the most important contributions of assessment to early childhood education is its ability to predict later achievement or school success (Tramontana et al., 1988). The typical method to evaluate the ability of an early assessment procedure to predict later achievement is through a pre-post design in which a sample of children are assessed early on and their later achievements after a time period are observed. The relationship between early assessment and later achievement measures is then examined and summarized statistically in the form of a correlation coefficient, which is often referred to as the predictive validity coefficient. This piece of statistical predictive evidence of validity has been suggested as the justification for early screening and intervention (Meisels, 1984; Stedman, 1981), for decision-making (Harber, 1981; Lewis, 1980), and for the development of programs (Beckman & Burke, 1984).

Although numerous validity studies have been conducted using this predictive validation design, the results appear to change in accordance with many testing conditions, including such test conditional variables as test type, developmental subconstructs being assessed, socioeconomic and demographic characteristics of the children assessed (e.g., SES, gender, income level, race), length of time for prediction, and the criterion outcome measures of later achievements used (Tramontana et al., 1988; Wilson & Reichmuth, 1985; Ysseldyke & O'Sullivan, 1987). Since predictive validity coefficients of early assessments are potentially influenced by so many possible variables, it is not easy to draw a generalized conclusion regarding the utility of early assessment in predicting later achievement or success. In other words, because each study has its own assessment and criterion measures, as well as other testing conditions, the observed validity coefficients may be interpretable only within unique circumstances. This problem leads to a very important theoretical and practical question: How can we make decisions regarding selection and usage of appropriate assessment instruments when faced with the inconsistent results of numerous validity studies?

With regard to the general predictive abilities of early assessments, two explanations can be addressed. First, it is possible that predictive validity is unique to each early assessment procedure and unique to each specific set of local testing conditions. For instance, when a standardized assessment instrument is utilized, its predictability to later achievement can be affected by characteristics of sample (e.g., gender, minority ratio, parent background), length of prediction (e.g., kindergarten-first grade, kindergarten-second grade), characteristics of criterion measure (e.g., reading, mathematics, social skills), and so forth. As a result, predictive coefficients using the same instrument can show variability across many different validity studies. If this is the case, no generalized statement can be made regarding the ability of early assessment to predict later achievement; rather, predictive ability will be different from situation to situation. Alternatively, the second possible explanation is that the variability in validity coefficients can be entirely attributed to a number of statistical artifacts (Schmidt & Hunter,