

THE EARLY ASSESSMENT CONUNDRUM: LESSONS FROM THE PAST, IMPLICATIONS FOR THE FUTURE

CATHERINE M. BORDIGNON AND TONY C.M. LAM

Ontario Institute for Studies in Education of the University of Toronto

The early childhood educational field has garnered attention with initiatives to foster skill acquisition in young children prior to kindergarten entry. These initiatives, in conjunction with the rigorous demands of curricular reform and a burgeoning accountability movement, invoke questions regarding the adequacy of the instruments used to assess young children and the inherent difficulties in conducting such assessments. Because the effectiveness of education relies critically on the sound diagnoses of children's readiness for learning and the measurement of their subsequent progression throughout the schooling process, critical issues in early assessment must be addressed. An examination of past practices was synthesized with recent research to focus awareness on the insufficient content domain, restrictive context, adverse timing and questionable psychometric properties, specifically the inappropriate norms and low predictive validity, of many instruments. Both the implications of and compensatory strategies for each issue are considered. © 2004 Wiley Periodicals, Inc.

Due to the impact of early childhood experiences, international initiatives have been developed to foster the cognitive, behavioral, and physical development of young children (McCain & Mustard, 1999). The current fervent interest in early education is propelled by two forces: a renewed recognition that early childhood experiences impact upon a child's self-esteem, social interactions and learning habits (Crahay, 1994; McCain & Mustard, 1999; Ontario Royal Commission on Learning, 1994), and the trend to raise standards of student achievement in an era of curricular reform and educational accountability (Bredekamp & Shepard, 1989; Shepard, 1997). Throughout North America, the curricular reform movement has led the impetus for earlier induction in mathematical and literacy skills, wherein curricular guidelines introduce cognitive skills at earlier stages than previously in a progression that is increasingly "... significantly more rigorous and demanding than previous curricula" (Ontario Ministry of Education and Training (MET), 1997, p. 3). This downloading of curricular skills to earlier grades and the escalation of curricular requirements and accountability pressures commences in the kindergarten years (Agostin & Bain, 1997; Bredekamp & Shepard, 1989; Meisels, 1995; Reichenauer, 1996; Shepard, 1997) and has been critiqued, in part for its inattention to the developmental patterns of children (National Association for the Education of Young Children (NAEYC), 1988, 1995; Shepard, 1997).

In concert with the more rigorous curriculum introduced at an earlier age has been the recognition that not all children are immersed in preschool experiences which foster creativity, knowledge acquisition, and acceptable behavioral practices (Crahay, 1994; Doherty, 1992; McCain & Mustard, 1999; NAEYC, 1995). Realizing that not all children arrive at the school door with equitable skills, the United States, in the late 1990s, proclaimed as its first National Educational Goal that by the year 2000, all children should start school ready to learn (Meisels, 1995). The discrepancy in early learning has been addressed in various localities by governmentally funded programs focusing on the care of preschool children. In the province of Ontario, the government has opened Early Years Centres across the province to highlight relevant aspects of child development and to enable caregivers to interact with early childhood professionals. The United States government, through its initiative, *Good Start, Grow Smart* (White House, 2002) aims to foster the early learning of young children, especially in the domains of prereading and language competency.

Correspondence to: Tony C.M. Lam, Ontario Institute for Studies in Education, University of Toronto (OISE/UT), Toronto, Ontario Canada M5S 1V6. E-mail: tlam@oise.utoronto.ca

To ensure that neither physical nor cognitive attributes hinder the attainment of educational foundation skills, an accurate assessment of a child's competency may be required. Appl (2000) cites the four major reasons for assessing preschool children: (a) to identify those children requiring further evaluation, (b) to provide diagnostic evaluation, (c) to assist in program planning, and (d) to monitor the progress of identified students. These reasons have a compelling focus in North America. Under the auspices of the United States' *Child Find* program (Appl, 2000), states are mandated to locate, identify, and service children with identified learning disabilities. As the intervention techniques and funding for such programs are contingent upon an accurate assessment process, it is imperative that the instruments used in this process are valid and reliable. The desirability for psychometrically sound instruments is of heightened importance given the requirements of the *No Child Left Behind Act* (U.S. Department of Education, 2002). Under the provisions of this act, a school is to be evaluated on the basis of its success in meeting designated assessment standards, on a yearly basis. The inclusion of children with disabilities in this evaluation is a compelling reason for the early identification of and subsequent programming for children who may experience difficulty in the school environment. During the formulation stage of Ontario's Early Years Centres, the suggestion was proffered (McCain & Mustard, 1999) that young children's development be assessed upon school entry. As the realms encompassed by this proposed assessment transcend the conventional areas of cognitive, language, and physical development to incorporate general knowledge and emotional well-being, the administration of traditional readiness or screening instruments would be insufficient.

The success of the aforementioned initiatives is predicated upon effective readiness and screening assessment practices which elicit information on domains of development. Research assessing traditional and emergent instruments enables practitioners to balance the merits, limitations, and risks of utilizing such instruments within an early assessment process. To facilitate this discussion in this paper, controversial issues, previously cited in the literature as hindrances in assessing preschool children validly and reliably, are reviewed. The implications of each issue, in conjunction with compensatory strategies, are discussed. Past studies are integrated with current research to inform the readiness and screening process. At this crucial juncture in the educational discipline, this discussion should propel requisite research and enable practitioners to more knowledgeably select, administer, and interpret appropriate assessment instruments.

ISSUES IN THE ASSESSMENT OF YOUNG CHILDREN

Previous research has indicated that approximately 10% of children have some form of learning disability (Kenny & Culbertson, 1993). To identify and program for those young children whose potential to succeed may be impacted by a disability, one or both of two types of instruments were traditionally used: (a) developmental screening, or (b) readiness tests. These instruments can, when valid and used appropriately, provide vital information about children's physical and cognitive status to parents and educational professionals. However, the administration of these traditional instruments has been fraught with charges of inappropriate construction, test misuse, and consequent invalidity (Bredenkamp & Shepard, 1989; Gredler, 1992; Meisels, 1987; 1994, 1995; Rafoth, 1997; Shepard & Graue, 1993).

The NAEYC (1995), in its position statement on school readiness, related three factors that are critical to the discourse on school readiness. Inextricably bound to this discussion are: (a) the diversity and inequity in children's early experiences, (b) the broad variation in their learning and developmental patterns, and (c) the extent to which the school-based kindergarten environment acknowledges and supports individual differences. These factors are entangled with the inherent difficulty in evaluating young children or their programs, especially the questionable suitability of commonly employed readiness and screening instruments and the unsuitability of young children

for testing purposes. Children, particularly those who have not yet experienced a formalized school setting, are adversely impacted by developmental variations (American Academy of Pediatrics, 1995; Bredekamp & Shepard, 1989; Farran & Shonkoff, 1994; NAEYC, 1990), experiential inequities (Bredekamp & Shepard, 1989; Gullo, 1991; Kagan, 2000; Lamberty & Crnic, 1994; Rafoth, 1997), and unfamiliarity with test taking (Shepard & Graue, 1993).

With accountability procedures and funding for early learning programs contingent upon children meeting designated standards, our ability to provide appropriate and effective assessments of young children cannot be disrupted by outmoded, inappropriate, and insufficient evidence. Otherwise, there is the danger that the realm of early childhood screening and readiness assessments may be not only ineffective, but impose negative and far-reaching consequences upon vulnerable members of society.

Gredler (1997) has proposed five criteria for judging readiness and screening instruments. The five characteristics of instruments that should be investigated are: (a) the standardization sample, (b) the cost of administration, (c) the ease of administration, (d) the content, and (e) the reported measures of reliability and validity. In concert with his recommendations and a systematic review of the academic literature and instruments of early assessment, five critical factors in the administration of readiness or screening instruments have been identified. These significant factors are an insufficient content domain, restrictive context, adverse timing, and inadequate psychometric properties, specifically inappropriate norms and low predictive validity. These factors are discussed and potential compensatory strategies are suggested.

Insufficient Content Domain

It is imperative that preschool assessment instruments tap valid indicators of subsequent ability to learn. What are these indicators and are they incorporated in commonly invoked instruments? The skills and developmental patterns that are the focus of many readiness and preschool instruments were those indicators viewed as important in the eras when the instruments were developed. The early childhood learning theories prominent at those times were the basis for the testing content and subsequent learning environment, neither of which are in vogue today (Bredekamp & Shepard, 1989; Gredler, 1992; Shepard, 1997; Shepard & Graue, 1993). Shepard (1997) citing research by Stallman and Pearson on readiness tests, critiqued the methodology employed by some readiness tests in which complex reading skills are decomposed into component tasks which are then tested in isolation. This procedure has been designated as outdated as reading acquisition is currently framed within whole language, constructivist (Mabry, 1995) or situated (Shepard & Graue, 1993) learning approaches. The content validity of the Metropolitan Readiness Tests has been faulted (Mabry, 1995; Stoner, 1995) due to its reliance on such an obsolete view of reading. The Gesell School Readiness Test was formulated from a maturational theory of development (Ilg, Ames, Haines, & Gillespie, 1978) and thus is in contradiction with present philosophies on child development (Gredler, 1992, 2000; Meisels, 1987; Shepard, 1997).

But with technological advances and early childhood day care programs enabling children to acquire vocabulary and expertise (such as in operating a computer) that were nonexistent 30 years ago, some aspects of readiness and screening instruments need to be reconfigured or replaced. In a meta-analysis review of more than 60 longitudinal studies comparing readiness measures and subsequent performance in the primary grades, La Paro and Pianta (2000) concluded that differences on school readiness screenings explain, at most, moderate variation on subsequent assessments. They further suggested that the realm of readiness screenings be enlarged beyond those of the traditional abilities and skills that are tested, to include some measure of social processes.

Crnic and Lamberty (1994) have cautioned that many readiness instruments do not tap domains that are strongly correlated with later cognitive demands nor is there necessarily a theoretical basis

for the inclusion of particular skills in screening and readiness instruments. Although valid preschool assessment is contingent upon integrating the precursors of later knowledge into the process, Lidz (1991) remarked that these precursors may actually be indistinguishable or too individualistic to allow for their incorporation in the assessment process. As we exist in a multicultural society, the determination of universal precursors may be an extremely difficult task. It is particularly noteworthy that the same construct may not be measured in an assessment process for minority and nonminority children (Barona, 1991).

Rafoth (1997) has observed that expertise in basic reading skills is essential for educational success. She therefore suggested that selected instruments assess early linguistic capacity, phonemic awareness, and verbal memory. However, in documenting the validation of the Preschool Learning Behaviors Scale, McDermott, Leigh, and Perry (2002) clarified the importance of teacher observation of children's learning behaviors, specifically the children's ". . . motivation, attention/persistence, strategy/flexibility, and attitudes toward learning" (p. 353). They contend that competence in learning behaviors is directly related to both academic and social attainment and should therefore be included in the assessment process.

Movements toward redefining readiness have also elicited calls that the scope extend beyond emergent numeracy and language skills to delve into knowledge pertinent to other curricular topics and to be reflective of the view that learning is multidimensional (Bredenkamp & Shepard, 1989; Ford & Harris, 1990). Readiness, as proclaimed by the United States National Education Goals Panel (Kagan, 1992, p. 50), has been hypothesized to be composed of five domains which are (a) physical well-being and motor development, (b) social and emotional development, (c) approaches toward learning, (d) language usage, and (e) cognition and general knowledge.

The NAEYC, in considering the realm of learning, has also acknowledged the affective component as integral to the learning process. This organization recognized that creativity, self-esteem, and social competence impact upon a child's progress in the school environment but also recognized the difficulty in assessing these characteristics (NAEYC, 1988). An inability to accurately assess social-emotional development may be linked to young children's sporadic and emotional behavior, in conjunction with problems in determining precursors of inappropriate, stable behavioral patterns (Lichtenstein & Ireton, 1984).

Although the recognition of potential perceptual, developmental, and cognitive deficiencies should be the primary focus of preschool assessment measures, attention needs to be focused as well on identifying social behaviors that may impede adjustment to the school environment (Agostin & Bain, 1997; Atkins-Burnett, Nicholson, & Meisels, 1997; Hoge & Wichmann, 1994; Wenner, 1995). To function successfully in our schooling system (and in our society) requires expertise in social as well as cognitive domains. Ontario's Royal Commission on Learning (1994) thus recommended a curriculum incorporating stages of cognitive and social development. Gullo (1994) reminded us to consider ". . . the multidimensional aspect of learning and development, as well as the multidimensional aspect of the environments in which they occur" (p. 23).

In sum, many readiness and screening instruments have been criticized (Crnic & Lamberty, 1994; Ellwein, Walsh, Eads, & Miller, 1991; Lewis, 1993) for their focus on narrowly defined content and isolated skills. However, Kenny & Culbertson (1993) noted that the developmental level, especially the limited language facility of young children, inhibits the provision of an extensive test content as does the restricted time frame for conducting the assessment (Bracken, 1988). These developmental limitations are not constrained to the presentation of test items or tasks but include restrictions on the manner and interpretation of responses (Gullo, 1997). Inappropriate responses from preschool children may not necessarily be indicative of a skill deficiency but may be attributed to impulsivity or avoidance behaviors. The effects of these limitations are compounded if the readiness or screening instrument serves as the only evaluation and is not an

element of a systematic, multifactor assessment process. The range of skills to be tapped and techniques to be used would invariably be constricted even further as younger children are included in the assessment process (Bracken, 1987; Mullen, 1992).

The inclusion of behavioral assessment, teacher observation, and caregiver information may modify charges of questionable reliability and validity garnered with the use of traditional screening and readiness instruments. Discrepancies in test content may be alleviated through the administration of systematic, multiple screenings (Kenny & Culbertson, 1993). As the range of skills and types of tasks appropriate for children expand as the children become older and immersed in educational activities, a more accurate assessment of knowledge and growth may be attained.

Restrictive Context

Meisels (1993, 1995) has stressed that since affective, cognitive, and psychomotor skill development are interrelated and occur in both the home and school environment, any assessment which is too divorced from the environment in which learning occurs, may be inaccurate. Thus, fellow educators (Bailey & Rouse, 1989; Lidz, 1991; Paget, 1989) have endorsed assessment in environments pertinent to the individual child's functioning, rather than a sole assessment with an unknown adult conducting the assessment. As children are observed in various environments, fluctuations in responses and behaviors can be contrasted to their performance during other phases of the assessment process.

A continuous, flexible assessment procedure would decrease the reliance on scores obtained from an artificial setting in which children may be less able to exhibit mastery of skills. Monitoring of children in various settings enables a contrasting of performance and the inclusion of a wider span of behaviors during the assessment process (Bracken, 1991). To veer from an artificial assessment context, recommendations are to incorporate play activities within the assessment process (Ontario Ministry of Education & Training, 1997; Ontario Royal Commission on Learning, 1994; Shepard, Kagan, & Wurtz, 1998). Young children's activities during play have been documented as contributing to their development of literacy, problem solving and social competence (Ontario Ministry of Education & Training, 1997) and in enabling teachers to plan activities to enhance this development (Meisels, 1995; Shepard, Kagan, & Wurtz, 1998). However, McDermott et al. (2002) aptly caution that there are not clear indications as to the means by which assessments conducted in nonformal environs convert into learning in the formalized school environment.

The diverse aspects of the readiness for learning and development of young children preclude a comprehensive picture of the child emerging from a single evaluation based on a solitary instrument. The limited attention span, sporadic behavioral patterns, and restricted tasks of which young children are capable, necessitate evaluation on more than one occasion and with differing modes of evaluation.

To provide a comprehensive portrait of the child's functioning, information should be gathered from multiple sources. A child's capacity to learn is influenced by innate ability, physical development, environmental stimulation and school programming (personnel, curricular structure and the availability of resources). Including information from both teachers and parents in the readiness or screening process has been recommended as a means to ensure that a broad assessment of the child is obtained and that the predictive validity of the screening is increased (Henderson & Meisels, 1994; Lamberty & Crnic, 1994; Lewis, 1993; Meisels, Henderson, Liaw, Browning, & Have, 1993; Williams, Gridley, & Treloar, 1989). The subsequent curricular process is also enriched by the unique information distilled from the triad of parent, child, and teacher (Gullo, 1994; Henderson & Meisels, 1994; Lewis, 1993; Meisels et al., 1993; Williams et al., 1989). However, alternative assessment procedures, even when serving as a supplement to standardized

procedures, need to be enacted with caution (Appl, 2000) until research has determined the viability and reliability of their components.

Adverse Timing

The realm of standardized and nonstandardized assessment has undergone considerable scrutiny in the past few decades. A focal point of discussion has been the determination of an appropriate time to commence a readiness, developmental screening, or diagnostic assessment process. When the testing situation occurs in the early days of kindergarten or preschool, poor performance may be tied to experiential inequities and aligned with the child's exposure to songs, riddles, stories, and other interactions that are prevalent in an enriched preschool environment. If a child's evaluated deficiencies on a readiness assessment or diagnostic instrument following screening are actually due to insufficient exposure or stimulation, inaccurate curricular provisions or remediation may serve to further disadvantage the child. However, beneficial results from early intervention have been cited (Agostin & Bain, 1997; Gullo, 1991; Keating, 1995; Lewis, 1993; Lichtenstein & Ireton, 1984; Nutall, 1992; Thurlow, 1992; Wolery, 1989) as the reason for testing early in the process to determine those children in need of remediation or support in the realms of affective, cognitive, and psychomotor domains. Countering this desire is the realization that low scores also may be attributable to initial lack of experience, shyness, anxiety, or inattentiveness in the school situation (Bagley, 1995; Bracken, 1991). The method of continual assessment if enacted with diligence, can identify children with potential learning difficulties at various stages of the assessment process. Nevertheless, practitioners need to temper the possible consequences of postponing diagnostic assessments for children who may experience learning or developmental hindrances, with the consumption of personnel time and funding allocations to assess children who will not require intervention.

Existent governmental policies pressure districts to locate and service children with disabilities as early as possible. Psychologists and program providers need to ensure that designations of deficiencies are therefore corroborated by other sources and alternative methods. Once children commence a program derived from the assessment, caregivers should be informed of their progress to ameliorate anxiety. A method of continual evaluation, in which a portfolio may be included, may improve the quality of young children's assessment and curricular progression.

Psychometric Properties

A premise of any testing situation, be it within a classroom setting, a local initiative or a large-scale assessment measure is that the assessment instrument has sound psychometric properties. Of particular importance in the context of assessing young children are the standardization sample (or norms) of the instrument, its reliability, and predictive validity (for screening instruments).

Inappropriate norms. For many instruments, the norming population was primarily composed of white, middle-class children. It is unlikely therefore that tests normed on this particular population would be representative of the diverse cultures in this era's early childhood learning environment. Use of inadequately normed instruments may thus have a deleterious effect on those children raised in cultures diverse from the normed population and those for whom English is not the primary language (Barona, 1991; Bracken, 1988; Bredekamp & Shepard, 1989; Farran & Shonkoff, 1994; Ford & Harris, 1990; Lamberty & Crnic, 1994; Rafoth, 1997).

Two separate studies of kindergarten screening practices in the state of New York have revealed the rather widespread usage of psychometrically questionable instruments. In their survey of screening practices, May and Kundert (1992) discerned that the Gesell School Readiness Test, the Developmental Indicators for the Assessment of Learning-Revised (DIAL-R), and the Brigance K &

1 Screen were administered in more than 10% of responding districts, a condition replicated in a similar study by Costenbader, Rohrer, and Difonzo (2000). The Gesell School Readiness Test has been faulted for outdated norms (Bradley, 1985) that are constrained by ethnicity, geographic area, and social class. The test has also been shown to have substantial variance in common with standardized intelligence measures and unacceptable test-retest reliability (Lichtenstein, 1990). Criticism of inadequate norms also extends to the Gesell Preschool Tests (Kaufman, 1989; Meisels, 1987). The Brigance K & 1 Screen has been disparaged for its failure to provide normative information (Berk, 1995; Shepard & Graue, 1993) and its lack of reliability and validity data (Berk, 1995; Costenbader et al., 2000; Shepard & Graue, 1993; Watson, 1995). Since the DIAL-3's normative sample replicated the 1994 United States census, this version of the screening instrument has been viewed as an improvement over earlier versions of the instrument (Fairbank, 2001), especially in the administration of the instrument to English-speaking children (Cizek, 2001; Fairbank, 2001).

To facilitate the development of psychometrically sound assessment instruments, both the norms and language incorporated in the instrument should be representative of the culturally diverse groups to be served (Lamberty & Crnic, 1994). This representativeness should also extend to an awareness of the cultural traditions in social interactions and behaviors. It has been conjectured (Bailey, 1989; Bracken, 1988; Gillespie-Silver & Scarpati, 1992; Meisels, 1994) that some tests and their norms may thus have penalized those children raised in cultures differing from the majority. Carney and Merrell (2002) further advocated for the development and use of assessment instruments in the primary language of the child's caregivers. They related that such an instrument would assist in ". . . foraging a meaningful collaboration between school psychologist, parents/guardians, and educators intervening on the child's behalf" (p. 367).

Instruments with more relevant norms have been introduced though these instruments have not yet received widespread usage nor are they appropriate for all situations. McDermott et al. (2002) have presented information on the standardization and validation of the Preschool Learning Behaviors Scale, an instrument that may assist in the planning of learning intervention programs for young children. Their normative sample corresponded to the 1992 United States preschool population, including stratification for ethnicity, parental education, and geographic area. Further research is planned to elucidate how assessment in preschool is coupled with learning in the formal school environment. Although not a new instrument, the Early Screening Inventory-Revised has been praised as an exemplary instrument for developmental screening (Paget, 2001) with an expanded normative sample and strong predictive validity (Kimmel, 2001; Paget, 2001).

Many instruments, whether designed for readiness testing, diagnostic evaluation, or developmental screening and regardless of their date of development, are further constrained as to their usefulness in that they may not have been standardized on children with severe handicaps (Bailey, 1989; DeMers, Fiorello, & Langer, 1992; Huang, Hunter, Reinert, & Wishon, 1992). With many developers attempting to formulate tests appealing to a wide-ranging population base, the inclusion of tests for specific populations may rely on a small norming sample or institutionalized group (Moore & McLaughlin, 1992). In selecting an assessment instrument, the standardization sample should mirror characteristics of the individual to be assessed. Otherwise, the application of such tests may lead to an erroneous determination of skill level unless corroborating evidence is obtained from multiple sources.

Low predictive validity. Prospective users are advised to ensure that all early childhood educational instruments have documented reliability in addition to the relevant type of validity (predictive for developmental screening test and construct for readiness measures). Invariably, commonly used readiness, screening, and behavioral instruments have been criticized on the grounds

of questionable validity, reliability, or both (Barona, 1991; Bracken, 1987; Bredekamp & Shepard, 1989; Carney & Merrell, 2002; Crnic & Lamberty, 1994; Harrison, 1991; Meisels, 1987; Rafoth, 1997).

It is imperative that an effective screening test be able to designate those children prone to develop difficulties later in school from those who will not. Kingslake's investigations (as cited in Gredler, 1992) have led to the recommendation that preschool screening tests should identify at least 75% of those children who would be unsuccessful in school. However, the high predictive error rate of many developmental screening instruments has been clearly cited (Costenbader et al., 2000; Graue & Shepard, 1989; Gredler, 1992, 1997, 2000; Shepard, 1992; Wenner, 1995). In reporting the results of studies documenting the predictive values of the DIAL-R, Brigance K & 1 Screen and the Early Screening Inventory (the precursor of the previously mentioned Early Screening Inventory-Revised), Gredler (1997) noted values in the range of .24 to .56. The error rate may be attributed, in part, to the insufficient content domain, as previously noted, of many early childhood assessment instruments. Generalizability is hampered since the instruments administered for documenting later performance varied among the studies, as did the ages at which the instruments were administered. Nevertheless, the inability to accurately discern children requiring intervention engenders a misuse of funding allocations and personnel support, as well as misguided caregiver concern.

Meisels et al. (1993), however, have stipulated that when conducting developmental screening an error of overreferral is generally less harmful than an error of underreferral. In documenting the limited predictive accuracy of tests for preschool children, Thurlow (1992) has noted that the administration of ensuing procedures may be more precise, especially with the inclusion of other sources of information. Since the determination of the predictive validity of an instrument necessitates longitudinal tracking of children, it has frequently been disregarded (Rafoth, 1997). The measures of yearly progress of children under the provisions of the *No Child Left Behind Act* (U.S. Department of Education, 2002) may facilitate constructive and updated analyses of the predictive validity of traditional instruments and emergent measures. Research has shown that the predictive validity of an assessment is improved when caregivers and teachers are involved in the process (Henderson & Meisels, 1994; Lamberty & Crnic, 1994; Lewis, 1993; Meisels et al., 1993; Williams et al., 1989).

DISCUSSION

For the early educational readiness, screening, and learning process to be reliable, valid and appropriate assessment instruments must be chosen prudently with consideration to the particular situation and individual to be evaluated. This deliberation includes an examination of the content, norms, and validity of the instruments and an acknowledgement of their deficiencies. We echo the recommendation (Dworkin, 1989; Gullo, 1994; Lamberty & Crnic, 1994; Meisels & Wasik, 1990) that a continuous appraisal system be invoked. Under this system, the veracity of the information gleaned from the assessment process could be improved and one could ensure that programming for individual children is appropriately focused and that developmental delays that appear later are identified (Meisels & Wasik, 1990). Lamberty and Crnic (1994) citing discussion at a conference on school readiness advocated iterative assessments on multiple occasions to facilitate programs of early intervention that recognize but not penalize variability in child development. A system of continuous appraisal does not penalize variations in learning patterns yet acknowledges the accomplishments of the individual at intervals which are not fixed to school year entry, exit, or adequate yearly progress measures. Recognition of the differential rates of learning and development in the primary grades enables the school program to be adapted to the individual, thereby fostering an uninterrupted learning progression for all children (American Academy of Pediatrics, 1995;

Gestwicki, 1995). The inclusion of affective and behavioral components in tandem with knowledge and skill developments fosters a broader and more effective “developmental surveillance” (Dworkin, 1989, p. 620) to enhance child well being. This methodology also entails the training of early childhood educators in the assessment and monitoring of young children’s development beyond that currently done in preschool and formal educational settings.

An individual child’s assessment results should be included with the results derived from the predominant individuals (parents, caregivers, teachers, medical personnel) influencing the young child’s development. As previously noted, the integration of the perspectives of the varied personnel with the information procured from multiple methods can serve to minimize the effects of instruments that are psychometrically deficient.

It is acknowledged that this direction in the early childhood assessment field requires an infusion of monetary resources and human personnel when community resources and educational institutions are already strained. Any developmental screening assessment, be it encompassed within a surveillance program or instituted within a narrower evaluative frame, has to balance the risk of underreferral when information from multiple parties is sought and is contradictory.

The early childhood assessment field would be well served by the continued introduction of innovative, judiciously designed instruments and techniques. Rafoth (1997) and Shepard (1997) have praised Meisels’ Work Sampling System as an innovative continuous assessment instrument that taps multiple domains and displays acceptable reliability and predictive validity. Teacher observations of a child’s performance are documented through developmental checklists, portfolios illustrating the child’s work, and summary reports. As checklists and summary reports are to be completed at three junctures in the year, the system enables student progress to be profiled on a regular basis. A preliminary study (Meisels, Liaw, Dorfman, & Nelson, 1995) examining the trial edition of the kindergarten checklist produced high measures of internal consistency ($.87 \leq \alpha \leq .94$). Moderate to high levels of reliability ($r = .69$ and $r = .89$) were found between the administration of the checklist at two time periods. Predictive validity indices were high ($.67 \leq r \leq .76, p < .001$) when the first and second administrations of the checklists were compared to two individually administered instruments, one tapping various cognitive domains and the other social behavior indices. Information on the psychometric characteristics and content domains of the Preschool Learning Behaviors Scale as reported by McDermott et al. (2002) (high interrater and test-retest reliability) and the Preschool and Kindergarten Behavior Scales as detailed by Carney and Merrell (2002) (high internal consistency) and Canivez and Rains (2002) illustrate the potential applicability of these instruments for children in the preschool population. To formulate, validate and analyze new processes, requires an infusion of funding and a commitment that such funding will be sustained. Financial resources must be sufficient to cover research investigations and training for early childhood personnel. When the primary method of assessing children as they enter the school setting are locally developed measures (in the Costenbader et al., 2000 study of New York state, 30% of responding districts used locally developed instruments), one can only hope that those individuals who have the task of formulating these instruments have been immersed in the psychometric properties of test theory and of the developmental philosophy of children. Given limited financial and personnel resources, those individuals responsible for conducting or interpreting the assessment need to be versed in recommending remediation methods that would be appropriate to strengthen skills and ameliorate weaknesses in the children who have received the assessment.

The analysis of critical issues in early childhood assessment cannot be extricated from the responsibility of the educational system to provide developmentally and culturally appropriate programs for young children. As the NAEYC (1988) has clearly dictated, the prime consideration in test use should be the improvement of educational experiences. Citing inadequate and possibly

detrimental interventions, educators (Bredekamp & Shepard, 1989; Graue & Shepard, 1989; Kirst, 1991; Shepard, 1992) have criticized the practice of basing instructional decisions (transitional year classes, retention, delayed entry) on the results of the administration of screening and readiness tests. In discussing a recommendation by Ontario's Royal Commission on Learning that school settings should be accessible for three-year-olds, Keating (1995) states that early childhood education should be envisioned as ". . . support for healthy and competent child development rather than as a reduced age for introducing formal academic instruction" (p. 15).

CONCLUSION

An analysis of previous research studies has identified critical factors in the assessment of young children. Significant among these factors are the insufficient content domain, restrictive context, adverse timing, inappropriate norms, and unsatisfactory predictive validity of many assessment instruments. This discussion of critical issues in the assessment of young children should lead to reflection by educators, researchers, and parents. Despite heightened scrutiny of children's performance and widespread invocation of standard-based assessment, our aim should be to design an assessment process that is reflective of current learning theories and the vast range of early childhood development for our multicultural population base. With earlier school entry becoming more feasible and with in-depth curriculum reform extending beneath the reaches of kindergarten, school systems must balance a multitude of conflicting demands to integrate developmentally and scholastically appropriate curriculum and assessment for the children passing through their doors.

An assessment process that could foster the energy, creativity, and enthusiasm that children initially bring to the school, should be the goal of all personnel designing, administering, and employing such instruments. As educational funding becomes contingent upon adequate progress, personnel are compelled to examine assessment instruments and methods to determine the adequacy of the content domain and psychometric characteristics. Where resources necessitate the use of less than flawless instruments, educators are cautioned to incorporate information from multiple sources and to contemplate the ramifications of the instruments' deficiencies in their analyses. When the assessment process can be invoked over time in relevant contexts, an accurate, comprehensive representation of the child may emerge.

REFERENCES

- Agostin, T.M., & Bain, S.K. (1997). Predicting early school success with developmental and social screeners. *Psychology in the Schools, 34*, 219–228.
- American Academy of Pediatrics. (1995). The inappropriate use of school "readiness" tests. *Pediatrics, 95*, 437–438.
- Appl, D.J. (2000). Clarifying the preschool assessment process: Traditional practices and alternative approaches. *Early Childhood Education Journal, 27*, 219–225.
- Atkins-Burnett, S., Nicholson, J., & Meisels, S.J. (1997). Assessing social competence in early childhood. In B. Spodek & O.N. Saracho (Eds.), *Yearbook in early childhood education: Vol. 7. Issues in early childhood assessment and evaluation* (pp. 149–178). Columbia University, NY: Teachers College Press.
- Bagley, D.M. (1995). Supporting families during the transition into Kindergarten. *Day Care and Early Education, 22*, 24–27.
- Bailey, D.B. (1989). Assessment and its importance in early intervention. In D.B. Bailey, Jr. & M. Wolery (Eds.), *Assessing infants and preschoolers with handicaps* (pp. 1–21). Columbus, OH: Merrill Publishing Co.
- Bailey, D.B., Jr., & Rouse, T.L. (1989). Procedural considerations in assessing infants and preschoolers with handicaps. In D.B. Bailey, Jr. & M. Wolery (Eds.), *Assessing infants and preschoolers with handicaps* (pp. 47–63). Columbus, OH: Merrill Publishing Co.
- Barona, A. (1991). Assessment of multicultural preschool children. In B.A. Bracken (Ed.), *The psychoeducational assessment of young children* (pp. 379–391). Needham, MA: Allyn & Bacon.
- Berk, R.A. (1995). Review of the Brigance K and 1 Screen for Kindergarten and First Grade Children Revised. In J.C. Conoley & J.C. Impara (Eds.), *The twelfth mental measurements yearbook [Electronic version]*. Retrieved August 12, 2002, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>

- Bracken, B.A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment*, 4, 313–326.
- Bracken, B.A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology*, 26, 155–166.
- Bracken, B.A. (1991). The clinical observation of preschool assessment behavior. In B.A. Bracken (Ed.), *The psychoeducational assessment of young children* (pp. 40–52). Needham, MA: Allyn & Bacon.
- Bradley, R.H. (1985). Review of the Gesell School Readiness Test. In J.V. Mitchell (Ed.), *The ninth mental measurements yearbook* (vol. 1, pp. 609–610). Lincoln, NE: Buros Institute of Mental Measurements.
- Bredenkamp, S., & Shepard, L. (1989). How best to protect children from inappropriate school expectations, practices and policies. *Young Children*, 44, 14–24.
- Canivez, G.L., & Rains, J.D. (2002). Construct validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales: Convergent and divergent evidence. *Psychology in the Schools*, 39, 621–633.
- Carney, A.G., & Merrell, K.W. (2002). Reliability and comparability of a Spanish-language form of the Preschool & Kindergarten Behavior Scales. *Psychology in the Schools*, 39, 367–373.
- Cizek, G.J. (2001). Test review of the Developmental Indicators for the Assessment of Learning, (3rd ed.). In B.S. Plake & J.C. Impara (Eds.), *The fourteenth mental measurements yearbook* [Electronic version]. Retrieved November 23, 2003, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Costenbader, V., Rohrer, A.M., & Difonzo, N. (2000). Kindergarten screening: A survey of current practice. *Psychology in the Schools*, 37, 323–332.
- Crahay, M. (1994). The IEA Preprimary Project: Obtaining knowledge to improve the quality of children's early experiences. In P.P. Olmsted & D.P. Weikart (Eds.), *The IEA preprimary study: Early childhood care and education in 11 countries* (pp.1–13). Tarrytown, NY: Pergamon.
- Crnk, K., & Lamberty, G. (1994). Reconsidering school readiness: Conceptual and applied perspectives. *Early Education and Development*, 5, 91–105.
- DeMers, S.T., Fiorello, C., & Langer, K.L. (1992). Legal and ethical issues in preschool assessment. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 43–54). Needham Heights, MA: Allyn & Bacon.
- Doherty, G. (1992, June). Addressing the issue of lack of school readiness in preschoolers. Ottawa, ON: Steering Group on Prosperity, Prosperity Initiative, Government of Canada.
- Dworkin, P.H. (1989). Developmental screening: Expecting the impossible? *Pediatrics*, 83, 619, 622.
- Ellwein, M.C., Walsh, D.J., Eads, G.M., & Miller, A. (1991). Using readiness tests to route Kindergarten students: The snarled intersection of psychometrics, policy, and practice. *Educational Evaluation and Policy Analysis*, 13, 159–175.
- Fairbank, D.W. (2001). Test review of the Developmental Indicators for the Assessment of Learning (3rd ed.). In B.S. Plake & J.C. Impara (Eds.), *The fourteenth mental measurements yearbook* [Electronic version]. Retrieved November 23, 2003, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Farran, D.C., & Shonkoff, J.P. (1994). Developmental disabilities and the concept of school readiness. *Early Education and Development*, 5, 141–151.
- Ford, D.Y., & Harris, J.J. (1990). On discovering the hidden treasure of gifted and talented black children. *Roeper Review*, 13, 27–32.
- Gestwicki, C. (1995). *Developmentally appropriate practice: Curriculum and development in early education*. Albany, NY: Delmar Publishers Inc.
- Gillespie-Silver, P., & Scarpati, S. (1992). Academic readiness in preschool children. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 233–245). Needham Heights, MA: Allyn & Bacon.
- Graue, M.E., & Shepard, L.A. (1989). Predictive validity of the Gesell School Readiness Tests. *Early Childhood Research Quarterly*, 4, 303–315.
- Gredler, G.R. (1992). *School readiness: Assessment and educational issues*. Brandon, VT: Clinical Psychology Publishing Company, Inc.
- Gredler, G.R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools*, 34, 99–106.
- Gredler, G.R. (2000). Early childhood education—Assessment and intervention: What the future holds. *Psychology in the Schools*, 37, 73–79.
- Gullo, D.F. (1991). The effects of gender, at-risk status, and number of years in pre-school on children's academic readiness. *Early Education and Development*, 2, 32–39.
- Gullo, D.F. (1994). *Understanding assessment and evaluation in early childhood education*. Columbia University, NY: Teachers College Press.

- Gullo, D.F. (1997). Assessing student learning through the analysis of pupil products. In B. Spodek & O.N. Saracho (Eds.), *Yearbook in early childhood education, Vol. 7: Issues in early childhood assessment and evaluation* (pp. 149–178). Columbia University, NY: Teachers College Press.
- Harrison, P.L. (1991). Assessment of adaptive behavior. In B.A. Bracken (Ed.), *The psychoeducational assessment of young children* (pp. 317–340). Needham, MA: Allyn & Bacon.
- Henderson, L.W., & Meisels, S.J. (1994). Parental involvement in the developmental screening of their young children: A multiple-source perspective. *Journal of Early Intervention, 18*, 141–154.
- Hoge, R.D., & Wichmann, C. (1994). An evaluative review of interview and rating/checklist instruments for assessing social/emotional competence in preschool children. Ottawa, ON: Department of National Health and Welfare.
- Huang, A.M., Hunter, L.R., Reinert, H.R., & Wishon, P.M. (1992). Assessment of children with mental retardation and other handicapping conditions. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 311–326). Needham Heights, MA: Allyn & Bacon.
- Ilg, F., Ames, L.B., Haines, J., & Gillespie, C. (1978). *School readiness. Behavior tests used at the Gesell Institute*. New York: Harper & Row.
- Kagan, S.L. (1992). Readiness, past, present, and future: Shaping the agenda. *Young Children, 48*, 48–53.
- Kagan, S.L. (2000). Making assessment count . . . What matters? *Young Children, 55*, 4–5.
- Kaufman, N.L. (1989). Test review of the Gesell Preschool Test. In J.V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook* [Electronic version]. Retrieved August 12, 2002, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Keating, D.P. (1995). Society's role in the development of young children. Weighing the research evidence. *Orbit, 26*, 14–16.
- Kenny, T.J., & Culbertson, J.L. (1993). Developmental screening for preschoolers. In J.L. Culbertson & D.J. Willis (Eds.), *Testing young children. A reference guide for developmental, psychoeducational, and psychosocial assessments* (pp. 73–100). Austin, TX: Pro-Ed.
- Kimmel, E. (2001). Test review of the Early Screening Inventory-Revised. In B.S. Plake & J.C. Impara (Eds.), *The fourteenth mental measurements yearbook* [Electronic version]. Retrieved August 12, 2002, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Kingslake, B.J. (1983). The predictive (in)accuracy of on-entry to school screening procedures when used to anticipate learning difficulties. *British Journal of Special Education, 10*, 24–26.
- Kirst, M.W. (1991). Interview on assessment with Lorrie Shepard. *Educational Researcher, 20*, 21–23, 27.
- Lamberty, G., & Crnic, K. (1994). School readiness conference: Recommendations. *Early Education and Development, 5*, 165–176.
- La Paro, K.M., & Pianta, R.C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*, 443–484.
- Lewis, E.G. (1993, January). Take a good look: Developmental screening. Paper presented at the Annual Conference and Exhibit of the Association for Supervision and Curriculum Development, Newton, MA.
- Lichenstein, R. (1990). Psychometric characteristics and appropriate use of the Gesell School Readiness Screening Test. *Early Childhood Research Quarterly, 5*, 359–378.
- Lichenstein, R., & Ireton, H. (1984). *Preschool screening: Identifying young children with developmental and educational problems*. Orlando, FL: Grune & Stratton.
- Lidz, C.S. (1991). Issues in the assessment of preschool children. In B.A. Bracken (Ed.), *The psychoeducational assessment of young children* (pp. 18–31). Needham, MA: Allyn & Bacon.
- Mabry, L. (1995). Review of the Metropolitan Readiness Tests (5th ed.). In J.C. Conoley & J.C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 611–612). Lincoln, NE: Buros Institute of Measurement.
- May, D.C., & Kundert, D.K. (1992). Kindergarten screenings in New York State: Tests, purposes, and recommendations. *Psychology in the Schools, 29*, 35–41.
- McCain, M., & Mustard, J.F. (1999). *Early years study*. Toronto, ON: Publications Ontario.
- McDermott, P.A., Leigh, N.M., & Perry, M.A. (2002). Development and validation of the Preschool Learning Behaviors Scale. *Psychology in the Schools, 39*, 353–365.
- Meisels, S.J. (1987). Use and abuses of developmental screening and school readiness testing. *Young Children, 42*, 4–6, 68–73.
- Meisels, S.J. (1993). Remaking classroom assessment with the Work Sampling System. *Young Children, 48*, 34–40.
- Meisels, S.J. (1994). *Developmental screening in early childhood: A guide*. Washington, DC: National Association for the Education of Young Children.
- Meisels, S.J. (1995). Out of the readiness maze. *Momentum, 26*, 18–22.
- Meisels, S.J., Henderson, L., Liaw, F., Browning, K., & Have, T.T. (1993). New evidence for the effectiveness of the Early Screening Inventory. *Early Childhood Research Quarterly, 8*, 327–346.

- Meisels, S.J., Liaw, F., Dorfman, A., & Nelson, R.F. (1995). The Work Sampling System: Reliability and validity of a performance assessment for young children. *Early Childhood Research Quarterly*, 10, 277–296.
- Meisels, S.J., & Wasik, B.A. (1990). Who should be served? Identifying children in need of early intervention. In S.J. Meisels & J.P. Shonkoff (Eds.), *Handbook of early intervention* (pp. 605–632). Cambridge: Cambridge University Press.
- Moore, M.S., & McLaughlin, L. (1992). Assessment of the preschool child with visual impairment. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 345–368). Needham Heights, MA: Allyn & Bacon.
- Mullen, Y. (1992). Assessment of the preschool child with hearing impairment. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 327–343). Needham Heights, MA: Allyn & Bacon.
- National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. *Young Children*, 43, 42–47.
- National Association for the Education of Young Children. (1990). NAEYC position statement on school readiness. *Young Children*, 46, 1–23.
- National Association for the Education of Young Children. (1995). *School Readiness. A position statement of the National Association for the Education of Young Children*. Washington, DC: Author.
- Nuttall, E.V. (1992). Introduction. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 1–7). Needham Heights, MA: Allyn & Bacon.
- Ontario Ministry of Education & Training (1997). *The Ontario curriculum. Grades 1–8. Mathematics*. Toronto, ON: Queen's Printer for Ontario.
- Ontario Royal Commission on Learning. (1994). *For the love of learning: Our vision for schools, Vol. II*. Toronto, ON: Queen's Printer for Ontario.
- Page, K.D. (1989). Assessment of cognitive skills in the preschool-aged child. In D.B. Bailey, Jr. & M. Wolery (Eds.), *Assessing infants and preschoolers with handicaps* (pp. 275–300). Columbus, OH: Merrill Publishing Co.
- Page, K.D. (2001). Test review of the Early Screening Inventory-Revised. In B.S. Plake & J.C. Impara (Eds.), *The fourteenth mental measurements yearbook [Electronic version]*. Retrieved August, 12, 2002, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Rafoth, M.A. (1997). Guidelines for developing screening programs. *Psychology in the Schools*, 34, 129–137.
- Reichenauer, L. (1996). High-Stakes kindergarten testing in Alberta? Some second thoughts. *Early Childhood Education*, 29, 10–12.
- Shepard, L.A. (1992). Psychometric properties of the Gesell Developmental Assessment: A critique. *Early Childhood Research Quarterly*, 7, 47–52.
- Shepard, L.A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools*, 34, 85–97.
- Shepard, L.A., & Graue, M.E. (1993). The morass of school readiness screening: Research on test use and test validity. In B. Spodek (Ed.), *Handbook of research on the education of young children* (pp. 293–305). New York: MacMillan.
- Shepard, L.A., Kagan, S.L., & Wurtz, E. (Eds.). (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.
- Stallman, A.C., & Pearson, P.D. (1990). Formal measures of early literacy. In L.M. Morrow & J.K. Smith (Eds.), *Assessment for instruction in early literacy*. Englewood Cliffs, N.J.: Prentice Hall.
- Stoner, G. (1995). Review of the Metropolitan Readiness Tests (5th ed.). In J.C. Conoley & J.C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 612–614). Lincoln, NE: Buros Institute of Measurement.
- Thurlow, M.L. (1992). Issues in the screening of preschool children. In E.V. Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers. Psychological and educational dimensions* (pp. 67–82). Needham Heights, MA: Allyn & Bacon.
- U.S. Department of Education, Office of Elementary & Secondary Education. (2002). *No Child Left Behind. A Desktop Reference*. Washington, DC: U.S. Government Printing Office.
- Watson, T.S. (1995). Review of the Brigance K and 1 Screen for Kindergarten and First Grade Children Revised. In J.C. Conoley & J.C. Impara (Eds.), *The twelfth mental measurements yearbook [Electronic version]*. Retrieved August 12, 2002, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Wenner, G. (1995). Kindergarten screens as tools for the early identification of children at risk for remediation or grade retention. *Psychology in the Schools*, 32, 249–254.
- White House (2002, April 2). *Good start, grow smart: The Bush administration's early childhood initiative*. Retrieved December 19, 2002, from <http://www.whitehouse.gov/infocus/earlychildhood/toc.html>
- Williams, K.T., Gridley, B.E., & Treloar, J.H. (1989). Using the parents' input in preschool screening. *Journal of Reading, Writing and Learning Disabilities International*, 5, 345–352.
- Wolery, M. (1989). Child Find and screening issues. In D.B. Bailey, Jr. & M. Wolery (Eds.), *Assessing infants and preschoolers with handicaps* (pp. 119–143). Columbus, OH: Merrill Publishing Co.

Copyright of *Psychology in the Schools* is the property of Jossey-Bass, A Registered Trademark of Wiley Periodicals, Inc., A Wiley Company and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.