Income Security & Social Development Studies
Applied Research Branch
Strategic Policy
Human Resources Development Canada

Études sur la sécurité du revenu et le développement social
Direction générale de la recherche appliquée
Politique stratégique
Développement des ressources humaines Canada

Tapping School Readiness in the NLSCY: Measurement Issues and Solutions

T-98-1E

by
Barbara A. Morrongiello

September 1997

The views expressed in Technical Papers are those of the authors and do not necessarily reflect the opinions of Human Resources Development Canada or of the federal government.

Les opinions exprimées dans les documents techniques sont celles des auteurs et ne reflètent pas nécessairement le point de vue de Développement des ressources humaines Canada ou du gouvernement fédéral.

■

The Technical Paper Series includes studies and research conducted under the auspices of the Applied Research Branch of Strategic Policy. Papers published in this series incorporate research of a technical nature intended for a limited number of specialists in particular areas of human resources studies.

La série des documents de travail comprend des études et des travaux de recherche réalisés sous l'égide de la Direction générale de la recherche appliquée, Politique stratégique. Il s'agit notamment de recherches analytiques techniques destinées à un nombre restreint de spécialistes dans des domaines particuliers liés aux ressources humaines.

■

First Edition/Première édition – Internet 1999

■

General enquiries regarding the documents published by the Applied Research Branch should be addressed to:

Publications Coodinator
Applied Research Branch
Strategic Policy
Human Resources Development Canada
140 Promenade du Portage IV, 4th Floor
Hull, Quebec, Canada
K1A 0J9

Telephone: (819) 994-3304
Facsimile: (819) 953-8584
E-mail: nancy.zalman@spg.org
http://www.hrdc-drhc.gc.ca/hrdc/corp/stratpol/

Si vous avez des questions concernant les documents publiés par la Direction générale de la recherche appliquée, veuillez communiquer avec :

Coordonnatrice des publications
Direction générale de la recherche appliquée
Politique stratégique
Développement des ressources humaines Canada
140, Promenade du Portage IV, 4e étage
Hull (Québec) Canada
K1A 0J9

Téléphone : (819) 994-3304
Télécopieur : (819) 953-8584
Courrier électronique : nancy.zalman@spg.org
http://www.hrdc-drhc.gc.ca/hrdc/corp/polstrat/

Abstract

This paper addresses measurement issues with respect to the interest of the National Longitudinal Survey of Children and Youth (NLSCY) in tapping school readiness in their evaluation of children in the 0 through 5 years age range.

The paper comprises four chapters. The first chapter considers the significance of school readiness for children's development and future well being, and review the five domains of functioning that have been shown to be relevant to children's school readiness (cf. Doherty, 1997). The second chapter discusses implications of findings about school readiness for measurement in the NLSCY, and discusses tests that are relevant to school readiness and that meet the administrative needs of the NLSCY (i.e., they are easy for respondents to understand, take little time to complete, and can be administered by individuals with minimal training). The third chapter discusses the strengths and weaknesses of existing measures with respect to school readiness and compares these with potential alternative measures. Finally, the fourth chapter concludes with some recommendations and justifications for measures that comply with the criteria for content development and which meet the procedural needs outlined by the NLSCY.

Résumé

Il est ci-après question des problèmes de mesure dans la perspective de l'intérêt des membres de l'équipe de l'Enquête longitudinale nationale auprès des jeunes et des enfants (ELNEJ) de vérifier la maturité scolaire dans le cadre de l'évaluation des enfants de 0 à 5 ans.

Le présent document comporte quatre chapitres. Le premier aborde l'importance de la maturité scolaire sur les plans du développement de l'enfant et de son bien-être dans l'avenir, et analyse les cinq domaines de fonctionnement qui, selon ce qui a été démontré, sont essentiels à la maturité scolaire de l'enfant (voir Doherty, 1997). Le deuxième chapitre traite de l'incidence des constatations relatives à la maturité scolaire aux fins de la mesure dans l'ELNEJ, et décrit les tests qui se rapportent à la maturité scolaire et qui satisfont aux besoins administratifs de l'ELNEJ (c.-à-d., tests qui sont faciles à comprendre par les répondants, qui prennent peu de temps à remplir et qui peuvent être administrés par des particuliers ayant un minimum de formation). Le troisième chapitre brosse un tableau des points forts et des points faibles des mesures existantes en ce qui concerne la maturité scolaire et compare ces dernières à des mesures possibles de remplacement. Enfin, le quatrième chapitre offre en guise de conclusion des recommandations et des justifications des mesures qui satisfont aux critères de l'élaboration du contenu et aux besoins en matière de procédures établis par l'équipe de l'ELNEJ.

Acknowledgements

Discussions with the following people aided in the preparation of this paper: Frank Mott (Centre for Human Resource Research, Ohio State University) and Dorothy Harris and Laurie Fowles (Thames Valley Children's Centre, London, Ontario).

Table of Contents

1. School Readiness

"School readiness" refers to a child's ability to meet the task demands of school and to acquire the curriculum content that is deemed appropriate for their grade at the time of entry into the school system (Kagan, 1992).

Readiness for school is a critical determinant of children's future well being for a number of reasons. First, research has demonstrated that school readiness at age six predicts children's academic success in the early school years (e.g., grade 3) and this, in turn, predicts the likelihood of children's successfully completing high school (Pulkkinen & Tremblay, 1992; Sanford, Offord, McLeod, Boyle, Byrne & Hall, 1994). Since employability in adulthood relates directly to level of formal education (Statistics Canada, 1996), young adults who fail to complete high school are at risk for having limited skills to compete successfully in the global market and for experiencing unemployment in adulthood (Ross & Shillington, 1990).

Second, lack of school readiness skills can lead to increased risk for developing peer problems. Lack of appreciation of socially acceptable behaviours and the inability to cooperate and express feelings in appropriate ways are likely to lead to adoption of less acceptable behaviour strategies in childhood, such as aggression and bullying, both of which are associated with peer rejection and isolation (Kupersmidt, Coie & Dodge, 1990; Ladd, 1990). Moreover, peer problems in the early school years typically persist throughout childhood and relate to interpersonal difficulties in adulthood (Kupersmidt et al., 1990).

Thus, school readiness has implications for a number of domains of functioning, and insufficient readiness can produce a number of long-term negative effects on children's well being and developmental outcomes.

In her paper entitled "Zero to Six - The Basis for School Readiness," Gillian Doherty (1997) elegantly discusses five components of school readiness. These include: physical well being and motor development; social knowledge and competence; emotional health and a positive approach to new experiences; language skills; and general knowledge and cognitive skills.

Physical well being and motor development

This includes sufficient *health* to minimize absences from school, adequate *nutrition* to provide for energy level, stamina, and concentration for the child to meet the demands of the school day (e.g., Worobey & Worobey, 1997), and age appropriate *motor coordination* skills for mastery of essential school tasks, such as printing and turning pages without ripping them (e.g., fine motor skills).

Social knowledge and competence

This includes an *awareness* of what are socially acceptable and age appropriate ways of relating to people (e.g., treating people with respect, being polite), as well as having the *communication* and *social skills* necessary to establish and promote meaningful relationships with others.

Emotional health and a positive approach to new experiences

This includes *emotional maturity* to delay gratification (e.g., turn taking), persist with repetitive classroom exercises such as phonics and counting, and regulation of one's emotions in order to handle failures and upsets in ways that are age appropriate and not disruptive or harmful to others. It also includes the capacity for reflection. Reflection before acting allows one to avoid acting too impulsively. Reflection on the consequences of one's actions allows one to learn from experiences.

Language skills

This includes having the skills to express one's ideas and feelings (*expressive* skills) and to understand what others are communicating verbally (*receptive* skills).

General knowledge and cognitive skills

This includes the ways in which children *organize* information, their capacity to *remember* information, their capacity to draw on prior knowledge in *assimilating new information*, and age appropriate *knowledge* about people, places, things, and events common to the child's living context.

2. Implications of School Readiness Findings for Measurement in the NLSCY and Evaluation of Some Relevant Measurement Instruments

A review of the literature on school readiness (e.g., Meisels, 1984) reveals three approaches to detecting children at risk for school problems: instruments that measure readiness directly, instruments that measure cognitive processing abilities that are relevant for school success, and instruments that measure developmental status.

Instruments that measure readiness directly

Tests of school readiness (e.g., revised Lollipop Test, revised Developmental Indicators for the Assessment of Learning test, DIAL-R) measure a child's degree of preparedness to benefit from preschool and kindergarten programs. Thus, readiness tests focus on children's current level of academic skills and behavioural competencies, and are often criterion-referenced tests, measuring skills kindergarten teachers deem essential for the child's success in kindergarten. Poor scores on these tests are presumed to reflect lack of experience or immaturity, rather than necessarily indicating underlying problems.

These tests yield a global score, with a cut-off score used to discern a child's state of readiness for school (i.e., pass or fail). Although some tests also yield subscale scores (e.g., DIAL-R), research indicates that it is best not to interpret these subscale scores individually because each yields a much lower classification reliability score in comparison to that obtained by using the global score (e.g., Suen, Czudnowski & Goldenberg, 1989). Thus, these tests provide a global measure of school readiness and are `screening' tests only; it is not appropriate to examine subscale scores individually for diagnostic purposes. In short, use of these tests may allow one to discern if a child is ready for school or not, but would not allow one to determine the bases for a child's lack of school readiness (e.g., academic, social-emotional, or communication abilities).

The most common tests of school readiness (Lollipop Test-Revised, Chew, 1989; DIAL-R test, Mardell & Goldenberg, 1975) that have proved somewhat useful as a measure of kindergarten

readiness and for predicting early school achievement (Chew & Lang, 1990; Eno & Woehlke, 1995) are discussed in greater detail in the Appendix.

Instruments that measure cognitive processing abilities that are relevant for school success

A second approach to school readiness, is to assess for more general cognitive abilities. Tests of cognitive processing skills (e.g., Mullen Scales of Early Learning) measure the child's abilities with respect to discrete processes that are the basis for children's learning, such as auditory and visual memory, attention, visual discrimination, auditory comprehension and object manipulation. In contrast to readiness and developmental tests, these tests do not usually yield a global score. In fact, it is the profile of scores across cognitive domains that one is most interested in with these tests, since this profile best illustrates discrete strengths and weaknesses relevant for educational planning whereas a global score would blend strength and deficit scores.

One drawback of these tests is their narrow focus (i.e., only on cognitive functioning). Certainly, recent research on school readiness factors (see Gillian Doherty's paper) highlight that there is a lot more to readiness for school than having the cognitive skills to meet academic demands of the setting. Social skills, emotional maturity, affect regulation and communicative competence are equally important determinants of school success. These tests do not tap any of these domains of readiness. Consequently, cognitive tests have more limited utility with respect to school readiness screening than tests that assess functioning in a broad range of domains, including social-emotional and communication areas.

Instruments that measure developmental status

Tests of developmental risk (e.g., Minnesota Child Development Inventory) identify children with current learning problems or delay or handicapping conditions. They assess children in a variety of domains (speech, language, gross and fine motor skills) and are designed to detect underlying impairments. Such impairments are likely to limit a child's ability to learn, but the tests themselves are not designed with this purpose in mind and are not marketed as tests of school readiness or achievement. Consistent with this, most are norm-referenced, as opposed to

criterion-referenced, tests. Nonetheless, these tests may prove quite useful for screening for school readiness for the following reasons.

These tests provide a global index of a child's developmental status, as well as reliable domain-specific scores that profile an individual's particular areas of strengths and weaknesses. Thus, these tests are useful for diagnostic purposes and for determining issues relevant for policy and program considerations. These tests also allow one to monitor changes in domain-specific functioning following interventions that target these domains. Finally, these tests allow for the greatest breadth of coverage in assessing the domains of functioning relevant for school readiness and outlined in Gillian Doherty's paper. Evidence is accumulating that school readiness is a multi-determined outcome that comprises a child's physical, social-emotional, and communication skills in addition to cognitive abilities. It seems likely therefore that developmental-based measures of school readiness afford the greatest potential for reliably assessing school readiness.

Since the NLSCY is interested in tapping school readiness skills in the context of gathering information on children's developmental status, primary consideration will be given in this paper to tests that index children's developmental status, either in general terms (i.e., yielding global scores) or in domain specific ways with emphasis on domains relevant to school readiness (e.g., motor development tests, language tests, etc), or both.

The findings on school readiness outlined in Dr. Doherty's paper have a number of implications for what the NLSCY needs to assess with respect to children 0 through 5 years of age. In the following section I consider each of the five readiness domains outlined in Chapter I and discuss “domain-specific tests” appropriate for assessing each aspect of school readiness. Following this, I consider instruments that yield a “general developmental index” and may be appropriate for tapping the various domains encompassed by school readiness. The sufficiency of the present measures with respect to school readiness is evaluated in the next chapter of the paper (Chapter III), with final recommendations on measurement given in Chapter IV.

2.1 Domain-specific Tests

Physical well being and motor development

The NLSCY needs to include measures that will provide an index of children's developmental status with respect to motor functioning. Although motor functioning encompasses both gross motor and fine motor skills, the demands of school place greater emphasis on fine motor functioning (e.g., holding and controlling a pencil, manipulating small objects, cutting with scissors) and research findings indicate that fine motor functioning discriminates behaviour problem and nonproblem preschool children better than gross motor, language, social and behavioural measures (e.g., Garrity & Servos, 1978). Consequently, an index of fine motor skills would seem especially important for assessing 4 and 5 year-olds with respect to school readiness in the NLSCY. At younger ages, a measure of gross motor functioning would likely serve to provide a sufficient index of developmental status for identifying children who are delayed in motor development, since gross motor functioning during the early months and years after birth reflects to a large degree neurological integrity (Gesell, 1973; Illingworth, 1975; Zaichowsky, Zaichowsky & Martinek, 1980).

Several tests of perceptual-motor functioning are available for consideration with respect to assessing fine motor skills relevant for school readiness, that is, motor performance skills that depend on visual-motor integration. Specifically, tests that assess the preschool child's ability to control a pencil and to copy symbols or figures are most germane to the classroom tasks that children face upon school entry (e.g., copying letters, numbers, shapes, etc) and seem most relevant for screening for school readiness (Simner, 1994). A review of several of the most commonly used tests for perceptual-motor assessment follow.

Bender Gestalt Visual Motor Test. This is likely the most popular of the visual-motor tests (Sattler, 1992). In this paper and pencil test children are asked to copy nine geometric figures (one at a time) onto a blank sheet of paper. Scoring is fairly straightforward (e.g., interrater reliabilities range from .79 to .99, Sattler, 1992) and involves scoring for accuracy based on degree of distortion, rotation, integration of parts of figures and degree of perseveration. This is a useful test for developing hypotheses about a child's perceptual-motor ability, although it is not sufficiently sensitive for diagnosis independent of other sources of information.

Four concerns have been written about the test: (1) percentile norms are available but only for children 5.0 years and older; (2) the test was poorly standardized (low N, sample was not representative of the general population, poor ethnic diversity), and (3) the test has only low to moderate (below -.40) correlations with measures of reading, arithmetic and school grades at the elementary grades (Blaha, Fawaz & Wallbrown, 1979; Caskey & Larson, 1980; Vance, Fuller, & Lester, 1986), indicating that the relationship between the test and academic skills is too weak to use the test *alone* to predict school readiness or subsequent achievement, although it has acceptable validity and reliability as a test of perceptual-motor functioning per se.

Beery Developmental Test of Visual-Motor Integration. This is a perceptual-motor test for children 4.0 - 13 years that was developed in 1964 and renormed in 1981. Reasonable reliabilities are reported in the manual: test-retest over 2 weeks to 7 months range from .63 to .92, inter-rater reliabilities range from .58 to .99. The concurrent validity reported in the manual is acceptable with respect to chronological age (r = .89) and perceptual skill (r = .80). Thus, the test is psychometrically sound.

For the purposes of the NLSCY, the test is easy to administer, requiring the child only to copy geometric forms that increase in difficulty, with testing discontinued after 3 consecutive failures and each design scored on a pass-fail basis. Although there are studies published illustrating the subjective nature of scoring judgements and that significant disagreements about scoring can result (Snyder, Snyder & Massong, 1981), there is high interrater reliability reported in the manual, and I have found the test easy to administer and score.

Denver Developmental Screening Test. This is a 105 item standardized test that is useful as a screening test for the detection of developmental delays in children 2 weeks to 6.4 years. This test taps a variety of domains of functioning relevant to school readiness, including motor functioning, with a distinction made between fine and gross motor competencies, and language and personal-social functioning. The test is easy to administer and score, and appropriate for repeated evaluations of the same child. The test is administered in about 10 minutes by asking the parent questions about the child and asking the child to perform various tasks.

Correlations ranging from .84 to .95 have been reported between this test and the Stanford Binet (L-M Form) and the Bayley Scales of Infant Development for 236 normal and mentally retarded

children (Frankenburg, Camp & Van Natta, 1971). Unfortunately, however, the standardization sample was extremely limited (Denver based sample only), the test is insufficiently sensitive in identifying delayed children under the age of three (Walker, Bonner & Milling, 1984), and, most importantly, the test has not proved effective for identifying children at risk for learning problems in the early school years (Diamond, 1990; Greer, Baucher & Zuckerman, 1989; Lindquist, 1982; Nugent, 1976). This test seems to do well in identifying children who are clearly delayed but is insufficiently sensitive for purposes of identifying preschool children who are marginally delayed but nonetheless in need of referral (Appelbaum, 1978; Diamond, 1990).

The Bruininks-Oseretsky Test of Motor Proficiency. This norm-referenced test is applicable for children 4 1/2 to 14 1/2 years of age, taps both gross motor and fine motor skills, and yields standard scores, percentile ranks and stanine scores. The full scale test takes 45 minutes to administer making it prohibitively long for the NLSCY. However, a 14 item form is available and provides a single score comparable to the overall Battery Composite score on the full battery. The short form takes about 15 minutes to complete, involves the child completing a few simple game-like tasks and requires no special training by the examiner. A number of investigations indicate that the short form is a reliable age-related measure for assessing motor proficiency among preschoolers (Beitel & Mead, 1980, 1982; Harrington, 1985). Moreover, discussions with field workers who have used the test in assessing for developmental delay in motor functioning at Thames Valley Children's Hospital in London, Ontario (personal communication, June 30, 1997) confirm the sensitivity and utility of the test (see Acknowledgements on page iv of this paper).

The standardization included Canadian and U.S. based children, and a stratified sampling procedure was used based on the 1970 US census data. Test-retest reliabilities range from .86 to .89 for the Battery Composite, and from .68 to .88 for the Fine and Gross Motor Composites. Construct validity tests indicate reasonably good correlations between subtest scores and chronological age (Mdn r = .78), and internal consistency measures reveal that the correlations between items and subtest scores are closer than between items and total test scores.

In summary, this is a psychometrically sound test that has proved useful for assessing preschool children's motor development status.

Social knowledge and competence

The NLSCY should incorporate measures that tap children's social knowledge and social competence. There are measures that tap this aspect of development, but do not do so in ways consistent with the needs of the NLSCY. For example, the Kohn Problem Checklist/Kohn Social Competence Scale assesses the social and emotional functioning of 3 to 6 year old children. The Checklist is an inventory of clinically significant problem behaviours and the Social Competence Scale measures the degree of competence with which a child functions within a school setting, yielding an Apathy-Withdrawal score and an Anger-Defiance score. However, these scales use the teacher as respondent, asking the teacher to report on the child's behaviour during the most recent week. A good alternative to a narrowly focused test on social competence per se, is the Vineland Adaptive Behaviour Scales. This test incorporates consideration of a number of domains of functioning that are relevant to social knowledge and social competence, as well it taps other domains relevant for school readiness using a parent survey format. This test is discussed at length under the section entitled "General Developmental Indicators."

Emotional health and a positive approach to new experiences

There are virtually no tests that address this aspect of development per se and do so in ways appropriate for administration in the NLSCY. Measures of infant's temperament come closest to assessing for a positive approach to new experiences, in that "difficult" temperament children are more often resistant to change and new experiences than are "easy" temperament infants (Thomas, Chess & Kohn, 1982). Information on temperament is presently taken in the interview portion of the NLSCY. Although such information does not allow one to differentiate typical from atypical emotional functioning, and should not be interpreted as an index of emotional health per se (i.e., children with difficult temperaments are not necessarily developmentally compromised in terms of emotional health), there is at least some coverage of this aspect of school readiness already included in the NLSCY and it is sufficient to allow one to differentiate children along one dimension critical for school readiness, namely -- adaptability to new experiences.

Language skills

Effective communication entails both expressive and receptive language skills, and there is sufficient evidence of *dissociation* between word comprehension and word production skills during childhood (e.g., Tomasselo & Mervis, 1994) that one should not infer normal language functioning on the basis of performance on assessment tests that tap only one of these domains of functioning. Rather, assessment of both aspects of language functioning are essential for a complete view of the developing child's linguistic and communicative status. With this in mind, the NLSCY's focus on receptive language alone is probably not providing sufficient sensitivity to assessing for language development problems.

A test that is administered and scored in ways comparable to the Peabody Picture Vocabulary Test-R (presently being administered to assess receptive language skills in the NLSCY) and assesses expressive language skills is the Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R; Gardner, 1990). This is a brief test that requires children to name each of a series of pictures (black and white line drawings) and is appropriate for children 2.0 years through 11 years 11 months. This is intended not just as a test for vocabulary, since many drawings require the abstraction of a concept (e.g., `musical instruments' drawing depicts several different types of instruments). Raw scores can be transformed through norm tables to obtain age equivalent scores, standard scores (Mean = 100, SD = 15), scaled scores (Mean = 10, SD = 3), percentile ranks and stanine scores.

The standardization sample is adequate in size at each age level, although it is poorly specified in the manual, making it unclear how representative the sample is of the broader population. Nonetheless, the psychometric properties are quite good, indicating a reliable and valid test for assessing expressive vocabulary. Internal consistency (i.e., split half) reliability is good (the median of the coefficients at each age was .90), although, test-retest stability is not reported. Validity estimates are based on correlating scaled scores on the EOWPVT-R with the Verbal Scale IQ on the Wechsler Preschool and Primary Scale of Intelligence-Revised (r = .73) and the Peabody Picture Vocabulary Test-Revised (r = .60). They indicate reasonable criterion related validity. Generally, most test critiques suggest that the test is best used as a screening test for expressive language vocabulary (Herman, 1994), as opposed to being used for drawing inferences about cognitive functioning per se. However, there is some evidence that it can

successfully predict IQ scores for young preschool children (e.g., Kutsick, Vance, Schwarting & West, 1988; Vance, West, & Kutsick, 1989), including those with mental retardation (Goldstein, Allen, & Fleming, 1982).

The MacArthur Communicative Development Inventories **(CDI).** This test (Fenson, et al., 1993) is a parent-completion inventory that takes about 20 - 30 minutes to complete and provides a general index of overall language competence as well as indices of more specific language skills (i.e., comprehension, production, gesturing, morphology and syntax).

This test is strong psychometrically, as evidenced by its demonstrated reliability and validity in a number of studies (e.g., Dale, Bates, Reznick, & Morisset, 1989; Dale, 1991; Camaioni, Caselli, Longobardi, & Volterra, 1991; O'Hanlon, Washkevich & Thal, 1991; Jackson-Maldonado, Marchman, Thal, Bates, & Gutierrez-Clellen, 1993). Internal consistency measures indicate good results for all three types of scales included in the test: the vocabulary scales demonstrate the highest internal consistency (alphas around .95), with very good results also for the grammatical scales (r = .80 to .91), and lower but adequate estimates for the gesture scales (r = .59 to .76). Good test-retest correlations have been obtained over an average interval of 1.35 months (r = .87, .95, and .86 for comprehension, production and gesture, respectively). Concurrent validity is quite good as evidenced by the high correlations between laboratory measures and inventory scores. For example, parental reports of children's vocabulary correlate with laboratory tests and free speech samples with coefficients ranging between .60 and .80 depending on the study (e.g., Dale, 1991). Similarly, laboratory based estimates of grammar correlate with performance on the grammatical complexity scale at r = .88 at age 1.8 years and r = .76 at age 2.0 years (Dale, 1991).

The CDI: Words and Gestures form is designed for use with 8-16 month olds, and is composed of two sections. The first part ("Early Words") includes a checklist of 396 words that are among the first to appear in the vocabularies of English speaking children. Next to each word the parent is asked to indicate if the child (a) understands the word, and (b) produces the word. The word checklist is divided into 19 broad categories (e.g., animal names, vehicle names, toys, food items, body parts, furniture, household objects). Part II ("Actions and Gestures") is a checklist of 63 communicative actions and/or symbolic gestures that also develop in this age range (e.g., pointing, nods head to indicate “yes”, blows kisses, plays peekaboo, uses imitative actions like “reading” a book).

The CDI: Words and Sentences form, which is for children between 16 and 30 months, is also composed of two sections. A vocabulary list (680 words organized into 22 categories) is presented in Part I ("Words Children Use") in which parents are asked only about production. Part II ("Sentences and Grammar") examines children's grammar from a variety of points of view (e.g., word combinations, irregular verb tenses) and asks parents to indicate how often the child uses a sentence structure (e.g., not yet, sometimes, often) and/or which of two utterances is most typical of how their child is presently talking (e.g., Kitty sleeping vs Kitty is sleeping). The format is easy to understand for respondents.

In summary, the CDI is a test that is easy to administer and score and is extremely cost-effective, providing tremendous yield in terms of in-depth language assessment for the short period of time involved in survey completion. There is no other test of language development that yields such a detailed assessment of a child's language in such a short period of time (Tomasello & Mervis, 1994).

General knowledge and cognitive skills

There are no tests that meet the administrative needs of the NLSCY and tap general knowledge and cognitive skills. However, some aspects of functioning in this domain are tapped in other measures already being taken by the NLSCY. Specifically, the PPVT-R provides an indirect index of cognitive level. Similarly, predictive validity tests of the MSD reveal that infant scores relate somewhat to PIAT Reading and PIAT Math scores 6 years later (Frank Mott, personal communication, June 1997). Thus, there is at least some coverage of this aspect of school readiness already addressed in the NLSCY measures.

2.2 General Development Indicators

An alternative approach to the selection of domain-specific measures to address the scope of school readiness issues and fill gaps in coverage, is to incorporate a parent report measure that broadly samples a child's developmental status. There are two tests that meet the administrative needs of the NLSCY: The Minnesota Child Development Inventory, and the Vineland Adaptive

Behaviour Scales. In the following section I review the Minnesota Child Development Inventory, discuss the appropriateness of this for school readiness screening, and indicate reasons that limit the appropriateness of the measure for adoption by the NLSCY. Following this, I review the Vineland Adaptive Behaviour Scales and discuss the appropriateness of this test for adoption by the NLSCY.

The ***Minnesota Child Development Inventory*** **(MCDI)** is a screening instrument for children that relies solely on the parents' report of children's behaviour and results in judgements that a child is appropriate, suspect, or inappropriate in development. The full inventory includes 320 items that describe behaviours of children 6 months to 6 1/2 years. These items are grouped into seven scales: Gross Motor, Fine Motor, Expressive Language, Comprehension-Conceptual, Situation Comprehension, Self Help and Personal-Social, each yielding age equivalent scores representing developmental levels in these distinct areas of functioning, with some of the scales shown to have some unique predictive validity when used individually (e.g., Chaffee, Cunningham, Secord-Gilbert, Elbard, & Richards, 1990; Dean & Steffen, 1984; Tomblin, Shonrock & Hardy, 1989). A summary scale, the General Development Index Scale (most appropriate for meeting the NLSCY administrative needs), includes 131 of the most age-discriminating items from the seven other scales to provide an overall index of development, the General Development Index (GDI). A child's development is *appropriate* if the score on the GDI is at or above the mean score for children 20% younger, *suspect* if the score falls in the range of children 21 to 30% younger, and *inappropriate* if the score falls below the mean score for children 30% younger.

Since the MCDI can take a long time to administer fully, an alternative to using the full version of the MCDI is to select certain scales from the MCDI. The scales that would cover the age range of interest to the NLSCY would be The Minnesota Infant Development Inventory for infants 0 through 15 months, the Early Child Development Inventory for children 6 months through 35 months, and the Preschool Development Inventory for children aged 36 though 47 months.

The ***Minnesota Infant Development Inventory*** **(MIDI)** includes items to measure development in five domains: gross motor, fine motor, language, comprehension, and personal-social. For each domain, one obtains an age-equivalent score. If the infant's score in a domain falls below the

average score for infants 30% younger, the infant is considered developmentally delayed. From the point of view of screening broadly for developmental delay, the utility of this inventory is that it yields domain-specific scores.

The ***Early Child Development Inventory*** **(ECDI)** includes six sections but to reduce length two are usually used: General Development (yielding a GDI score) and Possible Problems. The General Development scale includes items that tap seven domains of functioning: language comprehension, expressive language, gross motor, fine motor, self help, situation comprehension, and personal-social. However, individualized scores for each domain are not obtained when one uses the General Development Scale. In fact, the manual clearly indicates that the item composition on the General Development Scale results in greatest sensitivity with respect to identifying children with *general* developmental delays or with *language* problems.

With respect to scoring, a child is identified as delayed on the ECDI if their GDI score is lower than the average score for children who are 20% younger (not 30% as for the MIDI), with different threshold scores used for males and females over 18 months (chronological age).

The ***Preschool Development Inventory*** **(PDI)** includes a General Development scale, a Possible Problems List, and a Child Description component in which the parent indicates any special problems, questions, or concerns. The General Development scale includes 60 items that tap functioning in seven domains. Listed in order of thoroughness of coverage, these domains include: language comprehension, expressive language, fine motor, self help, personal-social, situation-comprehension, and gross motor. The General Development scale yields a GDI score but no breakdown by domain. As indicated in the manual, the test (like the ECDI) is most sensitive to children with *general* delays and *language* problems, is less sensitive to specific developmental problems, and is insensitive to gross motor problems and to social-emotional problems. If the parent reports 3 or more behaviour problems or 1 or more uncommon symptoms, the child is deemed to have a behaviour problem. If the child's GDI falls below the appropriate age and gender cutoffs (25% younger cutoff is used, not 30% as for the MIDI or 20% as for the ECDI) the child is deemed to be delayed. A child is identified as having a “disorder” if one of two conditions is met: the GDI score indicates delay or the child meets the definition for having a behaviour problem.

The psychometrics of the MCDI appear to be excellent. Reliability is high (Ireton & Thwing, 1974) and validity studies reveal high correlations between the MCDI and various standardized measures such as criterion measures for normal preschool age children (Gottfried, Guerin, Spencer & Meyer, 1983, 1984; Guerin & Gottfried, 1987) and children with developmental problems (Byrne, Backman & Smith, 1986) as well as for high risk infants (Saylor & Brandt, 1986). Moreover research on its clinical utility reveals good classificatory sensitivity (i.e., correct identification of children with delay = true positives) and specificity (i.e., correct identification of children without problems = true negatives). For example, Byrne et al. (1986) found that the MCDI correctly classified 83% of 71 preschool children, with a 97% sensitivity rate and a 73% specificity rate; similarly high estimates for sensitivity and specificity have been noted in other research with children 30 months and older (e.g., Gottfried, et al., 1984). Somewhat less accurate sensitivity and specificity estimates have been observed for infants under 30 months of age (Byrne et al., 1986; Shoemaker, Saylor & Erickson, 1993; Kopparthi, McDermott, Sheftel, Lenke, Getz & Frey, 1991) but, generally, the literature supports the predictive and concurrent validity of the MCDI with various populations.

Although the MCDI has proved to be an effective screening device for purposes of identifying children with developmental delay, the usefulness of the MCDI for predicting school readiness or subsequent failure has not been established. Indeed, in the research I could find in which this question was addressed, it has not proved to be the instrument of choice. For example, Schraeder (1993) compared the ability of three preschool risk-detection instruments to identify healthy very-low-birthweight children who would subsequently experience school failure. She found that the highest predictive accuracy and the fewest errors were obtained using an information processing test (Mullen Scales of Early Learning), followed by the instrument that used a developmental risk approach (GDI on the MCDI), with the least accurate outcomes resulting from the test that tapped kindergarten readiness (Minnesota Preschool Development Inventory). Children who subsequently experienced school failure and were not identified using the MCDI were most likely to be those with perceptual impairment or behavioural disturbances, suggesting that if the NLSCY were to incorporate the MCDI, this test would need to be further supplemented by a measure to assess visual-motor functioning and behavioural disturbances in order to improve predictive accuracy for purposes of tapping school readiness.

These findings, coupled with the fact that the General Development Inventory versions of the ECDI and PDI instruments, which are the inventories the NLSCY would have to use in order to keep administration time reasonable, do not yield domain specific scores, argue against adoption of the MCDI by the NLSCY.

The Vineland Adaptive Behavior Scales (VABS, Sparrow, Balla, & Cicchetti, 1984) was originally published as the Vineland Social Maturity Scale (Doll, 1965) but the current version is the one discussed herein since it has more recent norms and yields standard scores and more accurate percentile scores.

The VABS assesses personal and social adaptability of individuals from birth through 18 years 11 months; *adaptive behaviour* is defined as the performance of daily activities required for personal and social sufficiency. The VABS measures adaptive behaviour in four domains: Communication, Daily Living Skills, Socialization, and Motor Skills (including fine motor and gross motor skills). The combination of these four domains form the Adaptive Behaviour Composite score, although one also can compute domain specific scores.

The Communication domain samples receptive, expressive, and written communication skills. The Daily Living Skills domain evaluates personal living habits, domestic task performance, and behaviour in the community. The Socialization domain focuses on interactions with others, including play, use of free time, and sensitivity and responsibility to others. The Motor Skills evaluates gross motor and fine motor coordination. The Parent Survey form also includes a Maladaptive Behaviour domain that deals with undesirable behaviours that may interfere with adaptive behaviour.

The Parent Survey form includes a record booklet containing 297 items administered over a 20-45 minute interval; the older the child the more survey items that may apply and the longer the survey takes to complete. The interviewer uses the record booklet during the assessment to record item passes and failures; criteria for scoring are indicated and easy to understand. For each of the four primary domains (Communication, Daily Living, Socialization, Motor) one obtains a normalized standard score (Mean = 100, SD = 15).

The VABS is a well standardized test (1980 US census data; stratified sampling procedure), with separate norms provided for mentally retarded, emotionally disturbed, and physically

handicapped children and adults. It is also a psychometrically sound test, as indicated by the fact that the Survey form manual indicates an impressive number of reliability and validity studies. Internal consistency reliabilities for the domains range from .70 to .90. Interrater reliabilities are quite good, ranging from .86 to .95 using raw scores. Domain test-retest reliabilities are excellent, exceeding .80 in all domains. Construct validity was demonstrated in three ways: (1) mean raw scores increased with age for all four major domains, (2) factor analysis of the general Adaptive Behaviour Composite resulted in one significant factor, accounting for 55-70% of the variance at all yearly age levels, and (3) correlations of VABS scores with tests of intelligence yielded low but positive correlations, as predicted.

A short version of the Vineland, known as the Vineland Screener applies to children from birth through 18.11.30 years of age, and was developed for research purposes, although I could locate no published articles in which this instrument was used. The VAB-S derives from the longer test and is administered in the same way, that is, through a semi-structured interview. For children 0 through 5 years 11 months 30 days of age, the screener includes 15 items in each of four domains: (1) Communication (expressive and receptive), (2) Daily Living Skills, (3) Socialization, and (4) Motor; for children 6.0 years and older, the Motor domain is not included.

The items were selected on the basis of reliability, domain coverage, and strength of correlation with the total scales. There are four sets of items, one each for children 0 to 2-11-30 years, 3 to 5-11-30 years, 6 to 12-11-30 years, and 13 to 18-11-30 years. The test yields standard scores in each of the domains sampled and one composite score (the Adaptive Behavior Composite), and takes about 15-20 minutes to complete. Although I could find no documentation that the screener itself is reliable and valid, the high correlations between each of the domains on the screener and the full VABS (correlations between .87 and .98) would certainly indicate this to be the case. Thus, the VABS-S seems an excellent alternative to the full VABS if one wishes to minimize administration time without sacrificing sensitivity. Nonetheless, since I was unable to do so, the NLSCY team should confirm availability of the screener for use prior to making a decision to adopt this measure.

In summary, the VABS is a cost-effective tool for tapping a variety of domains of functioning relevant to school readiness, as well as providing a composite score on adaptive behaviour status.

3. Strengths and Weaknesses of Existing Measures

Reviewing the NLSCY cycle 2 questions and instruments reveals adequate coverage in some areas and shortcomings in other areas relevant to the five domains of school readiness discussed in Chapters I and II. I begin this chapter by reviewing the two primary instruments assessing child development that are currently being used in the NLSCY, namely-- The Peabody Picture Vocabulary Test-Revised (PPVT-R) and the Motor and Social Development Scale (MSD). Subsequently, I consider each of the five domains of school readiness outlined in Chapters I and II and discuss the adequacy of existing measures to tap functioning in each of these domains, including mention of questionnaire-based data being gathered by the NLSCY that might apply in each domain.

The Peabody Picture Vocabulary Test (PPVT-R). This is probably the most popular and the best test for assessing receptive language skills among children. It has excellent psychometric properties, and is a useful indicator of scholastic aptitude as an initial screening device (Umberger, 1985). In short, there is no better test than the PPVT-R for screening for receptive language skills. This was an excellent choice for inclusion in the NLSCY. This test should not, however, be assumed to provide an index of cognitive ability per se, since it is limited to what knowledge can be demonstrated via receptive language. In addition, it should not be interpreted as providing an index of school readiness, since there is no demonstration in the literature of a relationship between performance on this test and school readiness per se and a review of the test manual reveals that this was not a purpose for which the test was intended.

The Motor and Social Development Scale **(MSD*).*** This scale is purported to measure motor, social and cognitive development in infants from birth through 36 months of age. The scale was developed by Dr. Gail Poe of the National Center for Health Statistics for use in the Child Health Supplement of the 1981 National Health Interview Survey in the US. The items were derived from standard measures of child development (Bayley Scales of Infant Development, Gessell, Denver Developmental Screening Test). There are age ranges at which each item's developmental milestone is generally reached (based on US data), with the 50% cutoff age point most often used. Based on the child's chronological age, mothers answer the fifteen most

appropriate of the 48 motor and social items. Thus, it is easy to administer and score and to use for determination of a child's developmental status with respect to motor and social functioning.

Based on examination of some NLSCY data, the MSD apparently "tops out" (reaches ceiling) for children approaching three years, thereby not providing a sensitive test for these older children. This is not surprising, in light of the items included. Reviewing the items listed there seems to be much greater emphasis on motor than social items. In fact, of the 50 items listed, I would assign 6 of these to be tapping social functioning, 10 to be tapping cognitive/language, and the majority of items, 34, to be tapping motor functioning (predominantly gross motor, not fine motor, skills). Extending this point, I could find no mention of a factor analysis confirming that the MSD yields three factor scores, one for social, one for cognitive and another for motor functioning. Furthermore, since parents only complete 15 successive items, with the start point varying depending on their child's chronological age, a parent would likely answer at most 2-3 social items and 2-3 cognitive items. I would conclude therefore that the MSD provides a much better index of motor than social or cognitive functioning. Moreover, given the emphasis on gross motor skills I suspect it will be more sensitive to identify children who are seriously developmentally delayed than to screen those with more subtle delays. Certainly, for purposes of tapping cognitive, social and motor skills relevant to school readiness, the instrument seems to lack sufficient breadth and depth of coverage to do this satisfactorily.

In conversation with Frank Mott (June 24, 1997) about use of the MSD in the National Longitudinal Survey of Youth in the US, he indicated that similar problems arose in their survey data, that is, they found it was not as sensitive for identifying delay among older children as for infants and there was no evidence of differentiated factors (cognitive, motor, social) when a factor analysis was applied to the items. He also indicated, however, that tests of its external validity suggests it has at least some independent predictive value for school relevant behaviour and activities, with better predictive outcomes achieved for assessing children 12 months and under than for 2 and 3 year olds, although the results were statistically significant for all three age groups. Specifically, relating the MSD scores to scores 6 years later on the Child Behaviour Checklist revealed several associations with later behaviour: children who scored high on the MSD (%ile scores) showed less antisocial, less externalizing, less dependency and less hyperactivity behaviours six years later, with the hyperactivity correlation being the largest of these. Scores on the MSD also were associated with cognitive performance: MSD scores

correlated, six years later, with PIAT Reading and PIAT Math scores. One point that is worth noting, however, is that the size of all correlations was very low (e.g., .24 and less) despite their achieving statistical significance. Thus, statistical significance may not indicate clinically significant results given the small size of the correlations between the MSD and outcome measures.

It seems likely that the predictive value of the MSD stems from the test's sensitivity to discriminate infants and young children who are clearly developmentally delayed, as indicated largely by gross motor functioning. The NLSCY would not lose the predictive utility of the MSD if they substituted a test for the MSD that continued to tap gross motor functioning but also did a better job than the MSD of sampling fine motor and social functioning.

In the remainder of the chapter, I consider in greater detail the sufficiency of existing measures and questionnaire-based data for tapping each of the five school readiness domains outlined in Chapter I.

Physical well being and motor development

Some aspects of children's health are assessed in the NLSCY via questionnaire data. For example, Health Status questions (directed to parents of children 0 - 6 years of age) probe perceptual functioning relevant for school performance (vision, hearing, speech production), and Medical/Biological Questions (for parents of children 0-3 years of age) provide an index of prenatal environment (e.g., maternal smoking, drinking and drug use during pregnancy) which is relevant to physical well being. In addition, responses to Behaviour questions (for parents of children 0-3 years) provide a sense of the degree to which the child's eating and sleeping is routinized and predictable, which also bears on physical well being. Some composite score based on parental responses to questions in these three areas (Health Status, Medical/Biological, Behavior) could be used to develop an index of physical well being in the NLSCY, particularly for children 0 - 3 years of age.

For motor development, there is a question about mobility (gross motors status) and about pincer grasp (fine motor status) performance, although these obviously provide very limited information about motor status per se.

With respect to specific tests, for children 3 years and under, the MSD probably provides adequate coverage of gross motor functioning, but the issue of too low a ceiling for children approaching 3 years of age is a problem that limits the NLSCY's sensitivity to discriminate among children at these ages. This problem is realized in the predictive validity tests applied to the US longitudinal data (Frank Mott, personal communication, June 1997) in which they found better predictive validity for the MSD for infants 12 months and younger than for 2 and 3 year olds (see Acknowledgements on page iv of this paper). Furthermore, there is little information presently gathered with respect to motor functioning among preschoolers.

On balance, these findings suggest that the NLSCY would do well to at least supplement the MSD measure of motor performance with another measure that provides for better specificity, and equal sensitivity, across the entire age range, with respect to screening children for delays in motor development. The test also needs to provide for greater consideration of fine motor functioning and visual-motor integration skills, particularly in the 4-6 age range.

Given evidence indicating the importance of motor functioning, specifically fine motor skills and visual-motor functioning, for school achievement and as an indicator of school readiness (see Chapter II), the NLSCY should consider adding another measure to tap fine motor and/or visual-motor skills more directly among preschoolers. The addition of a more specific measure of motor functioning among preschoolers would certainly enrich the database with respect to assessing children's school readiness in the motor domain.

Social knowledge and competence

Some of the questionnaire data are relevant to children's social knowledge and competence. For example, some Behaviour questions for parents of children 2-3 and 4-11 years of age assess empathy (Q6D, Q6BB, Q8BB), helpfulness (Q6H, Q6U, Q6GG, Q6-SS, Q8D, Q8U), cooperativeness (Q6W, Q8W), and maturity of conflict resolution approaches (Q6J, Q6M) in dealing with peers and siblings. In addition, Relationship questions for parents of children 4 years and older provide an index of number of friends, and the parent's perception of the child's getting along with others (i.e., friends, teachers, parents, siblings) during the last 6 months. Finally, for 4-5 year olds, teachers are asked to provide an overall rating of a child's social-emotional development, relative to their age mates in junior or senior kindergarten, considering

such characteristics as adaptability, cooperation, interaction, responsibility and self control. In addition, teachers provide individual ratings about cooperation with other children, ability to follow instructions and rules, respect for others and their property, self control, and acceptance of responsibility for one's actions. Given the aspects of social competence and knowledge tapped via interviews, there is probably sufficient coverage in this domain of school readiness for 4-5 year olds, if one were to compute composite scores based on parent and teacher responses to questions about social maturity and developmental status. However, for children under 4, interview data does not provide a sufficiently broad sampling of social behaviour to index social knowledge and competence. Moreover, there would be little continuity in assessment between the younger (0-3 years) and older (4-5 years) ages if social development status was estimated in different ways for these different age groups.

With respect to formal tests, owing to the item composition, the MSD provides quite limited assessment of social functioning for children 3 years and under. Indeed, I think a case could be made that it provides virtually no index of social functioning at all (e.g., factor analysis of the items does *not* yield separate factors, including one for social). An additional measure of social functioning needs to be incorporated into the NLSCY to address this aspect of school readiness more thoroughly and directly across the entire 0 through 5 years age range.

Emotional health and a positive approach to new experiences

The temperament measures taken via interview in the NLSCY provide some coverage with respect to children's approaches to new experiences (e.g., “easy” and acceptable of change vs “difficult” and resistant to change.) In addition, some other interview questions provide indirect indicators of children's emotional health. For example, in the Behavioral questions for parents of children 4-11 years of age, some questions tap emotional maturity in the sense of assessing how children manage anger with peers (e.g., Q6R, Q6Z, Q6LL, Q6TT). However, in light of the importance of this domain for school readiness (see Chapters 1 and 2), there is probably insufficient coverage of this aspect of development in the NLSCY.

Language skills

Presently, the NLSCY includes a measure of receptive language vocabulary (PPVT-R). However, there is no direct measure of expressive language and there is little coverage of this aspect of child development on the present parent report inventories. Since the ability to express oneself verbally is essential for interpersonal relations and meeting classroom demands for expressing one's knowledge and communicating with others effectively (see Chapter II), this aspect of language functioning should be directly assessed in order to obtain a more complete picture of children's communicative competence in the NLSCY.

General knowledge and cognitive skills

The PPVT-R provides an indirect index of global cognitive functioning, but it does so in a very limited way (via vocabulary knowledge). Interview questions directed to the mother about the child's education (e.g., repetition of kindergarten) provide additional information about school success but no information on cognitive contributions, if any, to the child's school success, or lack thereof.

A test that samples adaptive behaviour may provide greater insights into general knowledge by assessing for daily living skills which draw on general knowledge about objects and events in the world (e.g., knowledge of the days of the week, understanding about currency and money, distinguishing safe from unsafe activities, recognizing the appropriateness of covering one's mouth when sneezing, etc). The NLSCY does not have adequate measures to tap general knowledge and could address this gap by incorporating a measure of adaptive behaviour functioning, since most adaptive behaviour scales assess the domain of daily living skills and these skills tap general knowledge about the world.

4. Recommendations and Justification for Measures Recommended

As the discussion in Chapter III indicates, there are some gaps in coverage that the NLSCY should address prior to the initiation of Phase 3 data collection. Specifically, measurement changes need to be made to achieve several desired outcomes, including: (1) better assessment of fine motor/visual-motor functioning among 4 and 5 year olds, (2) an index of expressive communicative competence at all ages, (3) greater breadth and depth of coverage of social development at all ages, (4) greater breadth and depth of coverage regarding emotional health at all ages, and (5) greater assessment of children's general knowledge about the world at all ages. In light of these five desired outcomes, three recommendations follow.

First, if the NLSCY feels that another 20-30 minutes of testing of children 4.0 and 5 years of age, and 10-15 minutes of testing of children 2 and 3 years of age, can be incorporated into their assessment protocol, then I would recommend two additions to their protocol to address the need for assessment of visual-motor functioning among preschoolers and expressive language at all ages:

- the Beery Test of Visual-Motor Integration for 4 and 5 year olds, and
- The Expressive One Word Vocabulary Test for children 2 through 5 years of age.

The Beery (as discussed in Chapter II) would provide a direct index of fine motor/visual-motor skills which are important for school success and are not presently being assessed. The Beery is easy to administer and score and can be completed in about 10-15 minutes. The advantage of the Beery over the Bender Gestalt is that is applies to a broader age range. The Beery applies to children from 4.0 years of age whereas the Bender Gestalt applies to children from 5.0 years of age.

The EOWVT (discussed in Chapter II) would yield an index of expressive vocabulary and, in conjunction with the PPVT-R which assesses receptive language, it would provide a more thorough assessment of language skills important for school readiness and achievement. The

EOWVT also can be completed in about 10-15 minutes. Although the CDI is a language test that provides a much greater wealth of information about communicative competence than the EOWVT, in fact, it probably provides more depth of analysis than is required by the NLSCY and, most problematically, it does not apply to preschoolers. Since language delays, if they exist, should start to be quite self evident during the preschool years, the NLSCY should include an expressive test for this age range. The EOWVT should provide sufficient sensitivity for the NLSCY to capture expressive language delays that might be relevant to school readiness.

Secondly, I recommend extending the interview with the mother by the 20 minutes necessary to complete the Vineland Adaptive Behaviour Scale (or the Screener). The VABS is very cost-effective, providing an excellent return of data for the time invested. It assesses functioning in a number of domains important for school readiness (expressive/receptive communication, social, daily living/general knowledge, fine/gross motor), and speaks directly to the NLSCY's need to achieve greater breath and depth of coverage with respect to social and emotional functioning and children's general knowledge. Moreover, it covers each domain with sufficient depth and breadth that one can get domain specific scores, as well as an overall score. I believe the administration guidelines could be adapted to meet the needs of the NLSCY and that some innovative technology would make training more manageable (e.g., developing a few training videotapes to illustrate how the interview should proceed for each domain- I have used this approach, with much success, to teach students to administer tests in the Child Assessment courses I teach).

Furthermore, since the VABS covers, to some extent, expressive language and fine motor skills, the NLSCY could forego adding the child measures recommended, *if* they felt adding these tests would make prohibitive demands of the child's time and attention or the examiner's assessment time. Since direct measures of child functioning are likely to be more sensitive to developmental status differences than parent reports on a child's level of functioning, however, this would not be an ideal outcome. Nonetheless, the fact is that the addition of the VABS-S, which offers both breadth and depth of coverage of four domains relevant to school readiness, will result in less urgency for the addition of further domain specific tests. Thus, incorporating the VABS alone

into the NLSCY protocol would likely be sufficient to produce significant improvements in coverage of school readiness without the addition of any further child measures.

Finally, with respect to relations between adaptive behaviour and school readiness, there is evidence that adaptive behaviour (as measured by the VABS) has significant effects on achievement beyond that accounted for by intelligence (e.g., Keith, Harrison & Ehly, 1986). Thus, the VABS is apparently tapping domains relevant to success in school.

Thirdly, I recommend not deleting the MSD from the NLSCY. Practically speaking, for the time it takes to administer these 15 items more is to be gained by continuing to include the measure than by deleting the measure, particularly since the VABS-S does not assess motor functioning at all. Since the MSD was used in prior phases of data collection and there is some evidence to indicate its utility in other longitudinal research (e.g., scores during infancy and early childhood relate to some aspects of cognitive functioning 6 years later in the National Longitudinal Survey of Youth in the US; Mott, 1997 - personal communication), it seems worthwhile to continue with the measure despite its limitations (see Chapter III). In addition, greater information about the sensitivity of the MSD is likely to be gained by being able to compare MSD and VABS-S scores directly in phase 3 data. Certainly, if the VABS-S proved a more sensitive index of identifying developmentally delayed children than the MSD in phase 3 this would be important information to have in interpreting the data from the MSD in phase 1 and 2 cycles. Finally, for purposes of comparing the status of children in Canada and the US, the usefulness of the MSD is self evident, since comparisons across countries are most appropriate when the same measures have been applied. For a variety of reasons therefore, I recommend continuing to include the MSD in the NLSCY.

In the following sections I elaborate in greater detail how the VABS-S would specifically address the five domains of school readiness outlined in Chapter II and of interest to the NLSCY and compare this with alternative measures, providing justification for selection of the VABS-S instead of the alternatives.

Physical well being and motor development

The Motor domains of the VABS-S consider both gross motor and fine motor skills, with appropriately increasing emphasis on fine motor skills assessment at increasing ages. Thus, the test would help fill the gap in assessing for visual-motor integration skills among preschoolers.

Although the Early Child Development Inventory (part of the Minnesota Child Development Inventory), also applies to children 3 years and under and incorporates some items to tap motor development, the item composition on the General Development Scale results in greatest sensitivity with respect to identifying children with general developmental delays or with language problems. There is no indication of the test's sensitivity with respect to screening for motor functioning problems per se. Similarly, for children 4 and 5 years of age, the General Development scales of the Preschool Development Inventory (part of the Minnesota Child Development Inventory) incorporates motor items. However, the manual indicates that the balance of items makes the General Development scale most sensitive to general developmental delays, as opposed to delays in any one domain, such as motor development. Thus, the VABS-S seems a better choice to fill the gaps evident in motor assessment.

Social knowledge and competence

There are a number of items in the Socialization Domain of the VABS for infants 36 months and under, which serve to illustrate a number of aspects of social functioning that are being missed with the MSD. Moreover, the Socialization domain provides a good index of children's adaptive behaviour with respect to social situations and interpersonal relationships, which speaks directly to the issue of children's social knowledge and competence (School Readiness point iii- see Chapter II); certainly, I could locate no other test that assessed child development in this domain and met the administrative needs of the NLSCY. Finally, the Daily Living Skills domain assesses the extent to which children are independent, self sufficient, and cooperative, which also are essential skills for school entry; again, I could locate no other test that assessed child development in this domain and met the administrative needs of the NLSCY. On balance, I believe, the VABS-S will provide coverage in the social domain, *and* yield information relevant to school readiness for which no other test could be located.

Emotional health and a positive approach to new experiences

Failure to include a direct measure of emotional health in the NLSCY may not be a problem, since the VABS-S includes items on the Socialization domain that tap aspects of children's emotional health (e.g., recognition of feelings such as anger, happiness, fear; recognition of other's feelings and needs). Thus, although the VABS-S does not yield a specific index of emotional maturity per se, the Socialization domain incorporates a consideration of this aspect of social-emotional functioning. Thus, this domain specific score could be used as a composite index of social development and emotional health (ie., considered as an index of these combined aspects of school readiness-- see Chapter II).

Language skills

The Communication domain of the VABS-S incorporates consideration of both expressive and receptive skills, and so provides for a broad sampling of communicative competence. Certainly, the broader one's sampling of communicative competence the more sensitivity the NLSCY will have to identify children with language problems that limit communication competence relevant for school readiness.

General knowledge and cognitive skills

The VABS-S includes items that tap a variety of aspects of cognitive functioning. For example, items listed under the Socialization domain tap awareness of cause-effect relations (e.g., Apologizes for unintentional mistakes), imitation and pretend play, and awareness of rules, all of which indicate memory for material. The Communication domain includes items that tap awareness of what would be considered humorous (e.g., tells jokes) and the Daily Living Skills domain includes items indicating awareness of danger, time, and ownership. Thus, limitations in cognitive functioning would likely be reflected in a lower than average composite score on the VABS-S. Certainly, this is what one would expect based on the fact that limitations in adaptive functioning is a *required* characteristic for assigning children a diagnosis of `mental retardation' (cf. Diagnostic and Statistic Manual of Mental Disorders IV, 1995), reflecting the fact that

adaptive behaviour and intelligence are separate, but related constructs (Keith, Fehrman, Harrison & Pottebaum, 1987; McMann & Barnett, 1984; Platt, Kamphaus, Cole, & Smith, 1991).

In conclusion, I recommend that, at a minimum, the NLSCY adopt the VABS-S for Phase 3. Adding the VABS-S fills gaps in coverage for which there are no alternative tests to be recommended. Furthermore, the incorporation of the VABS-S has the additional benefit of reducing the need for other measures (Beery, EOWVT) that tap functioning in specific domains relevant to school readiness but for which there is presently insufficient coverage in the NLSCY. The VABS-S is psychometrically sound and meets the administrative needs of the NLSCY. There is no obvious loss of sensitivity for screening for developmental delay using the VABS-S instead of the full scale VABS.

Appendix

School Readiness Tests

The two most popular tests are the Lollipop Test and the Developmental Indicators for the Assessment of Learning (DIAL-R) test.

The Lollipop Test: A Diagnostic Screening Test of School Readiness. This test (Chew, 1989) provides a criterion referenced measure of preschool and kindergarten readiness and comprises four subtests: (1) Identification of Colours and Shapes and Copying Shapes, (2) Picture Description, Positions and Spatial Recognition, (3) Identification of Numbers and Counting, and (4) Identification of Letters and Writing. The test requires the use of stimulus cards and individual interaction between pupil and examiner, takes about 15 minutes to administer and score, can be easily administered by laypersons with little training, and scores have been shown to be relatively independent of socio-economic variables (Chew & Morris, 1987). The difficulty level is pitched so that children who are ready for school will find it relatively easy to complete. Concurrent validity of the test has been established in that it has been shown to correlate .76 with scores on the Metropolitan Readiness Test (Chew & Morris, 1984) and .74 with scores on Word Reading and .72 with scores on the Stanford Achievement Test (Chew & Morris, 1989). Moreover, recent evidence reveals satisfactory predictive validity with respect to early school achievement (Chew & Lang, 1990; Chew & Morris, 1989), and suggests that the test yields a greater range of scores than the DIAL test thereby being less likely to produce problems of ceiling performance levels (Chew & Morris, 1987). One limitation of the test, however, is that no norms are available to indicate ages at which one should interpret failures as indicating delay. Rather, the manual suggests that local norms be developed and that the items failed be taken as indicating areas needing remediation. In summary, this is a short, easy to administer test that is psychometrically sound and serves the purpose of screening for school readiness.

The ***DIAL-R*** is probably the most popular and psychometrically sound screening test to identify pre-kindergarten children with potential learning problems or who are gifted (see Miller & Sprong, 1986 for comparison with other similar tests); the revised test is available from the American Guidance Service. It is an untimed, teamed administered, norm-referenced test that

Appendix (continued)

yields scaled scores and percentile ranks for (1) Motor (fine & gross), (2) Conceptual, and (3) Communication Skills. It can be administered to children between 2.0 and 6.0 years in about 20 minutes; children indicate their responses orally or by making a motor response. Administration requires the child to visit three areas or stations, each containing eight short tests. For example, the Motor area tasks include activities such as jumping, hopping, skipping, cutting with scissors, copying shapes and letters, and writing one's name.

The test was well standardized, reliability coefficients range from .76 to .90 (test-retest) and from .87 to .92 (alpha), and validity studies have yielded good sensitivity (.70) and specificity (.96) estimates (Mardell-Czudnowski, Goldenberg, Suen, & Fires, 1988). A number of studies also reveal good predictive validity (Chew & Lang, 1990; Jacob, Snider, & Wilson, 1988; Mardell & Goldenberg, 1975; Vacc, Vacc & Fogleman, 1987) and classificatory power (Cooper, & Shepard, 1992; Suen, Mardell, & Goldenberg, 1989). Although some researchers have suggested that subtest ceiling effects at older ages sometimes limit use of the test to predict school success (Chew & Morris, 1987; Obrzut, Bolocofsky, Heath & Jones, 1981), this multidimensional screening test has been widely adopted for use in screening programs throughout North America. The content is learning oriented and appears to be well suited for preschool screening.

References

Appelbaum, A. (1978) Validity of the Revised Denver Developmental Screening Test for referred and nonreferred samples. *Psychological Reports*, 43, 227-233.

Beitel, P., & Mead, B. (1980). Bruininks-Oseretsky Test of Motor Proficiency: A viable measure for 3- to 5-year old children. *Perceptual and Motor Skills*, 51, 919-923.

Beitel, P., & Mead, B. (1982). Bruininks-Oseretsky Test of Motor Proficiency: Further verification with 3- to 5-year old children. *Perceptual and Motor Skills*, 54, 268-270.

Blaha, J., Fawaz, N., & Wallbrown, F. (1979). Information processing components of Koppitz errors on the Bender Visual Motor Gestalt Test. *Journal of Clinical Psychology*, 35, 784-790.

Byrne, J., Backman, J. & Smith, I. (1986). Developmental assessment: The clinical use and validity of parental report. *Journal of Pediatric Psychology*, 11, 549-559.

Camaioni, L., Caselli, M., Longobardi, E., & Volterra, V (1991) A parent report instrument for early language assessment. *First Language*, 11, 345-359.

Caskey, W., & Larson, G. (1980). Scores on group and individually administered Bender Gestalt Test and Otis Lennon IQs of kindergarten children. *Perceptual and Motor Skills*, 50, 387-390.

Chaffee, C., Cunningham, C., Secord-Gilbert, M., Elbard H., & Richards, J. (1990). Screening effectiveness of the Minnesota Child Development Inventory Expressive and Receptive Language Scales: Sensitivity, Specificity, and Predictive Value. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 80-85

Chew, A.L. (1989*). The Lollipop Test: A Diagnostic Screening Test of School Readiness-Revised*. Atlanta, GA: Humanics Limited.

Chew, A.L., & Lang, S. (1990). Predicting academic achievement in kindergarten and first grade from prekindergarten scores on the Lollipop Test and DIAL. *Educational and Psychological Measurement*, 50, 431-437.

Cooper, D., & Shepard, K. (1992). Review of DIAL-R. *Learning Disabilities Research and Practice*, 7, 171-174.

Dale, P. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Sciences*, 34, 565-571.

Dale, P., Bates, E., Reznick, S., & Morisset, C. (1989). The validity of a parent report instrument of child language at 20 months. *Journal of Child Language*, 16, 239-249.

Dean, R. & Steffen, J. (1984). Direct and indirect pediatric screening measures. *Journal of Pediatric Psychology*, 9, 65-75.

Diagnostic and Statistic Manual of Mental Disorders (IV ed.), Washington, DC: American Psychiatric Association.

Diamond, K. (1990). Effectiveness of the Revised Denver Developmental Screening Test in identifying children at risk for learning problems. *Journal of Educational Research*, 83, 152-157.

Doll, E. (1965). *Vineland Social Maturity Scale*. Circle Pines, MN: American Guidance Services.

Eno, L., & Woehlke, P. (1995). Use of the Lollipop Test as a predictor of California Achievement Test scores in kindergarten and transitional first-grade status. Psychological Reports, 76, 145-146.

Fenson, L., Dale, P., Reznick, J.S., Thal, D., Bates, E., Hartung, J., Pethick, S. & Reilly, J.(1993). *The MacArthur Communicative Development Inventories: Users Guide and Technical Manual*. San Diego: Singular Publishing Group.

Frankenburg, W., Camp, B., & Van Natta, P. (1971). Reliability and stability of the Denver Developmental Screening Test. *Child Development*, 42, 1315-1325.

Gardner, M.F. (1990). *Manual for the Expressive One-Word Picture Vocabulary Test-Revised*. Novato, CA: Academic Therapy Publications.

Garrity, L. & Servos, A. (1978). Comparison of measures of adaptive behaviours in preschool children. *Journal of Consulting and Clinical Psychology*, 46, 288-293.

Gesell, A. (1973). *The first five years of life: A guide to the study of the preschool child*. New York: Harper & Row.

Goldstein, D., Allen, C., & Fleming, L. (1982) Relationship between the Expressive One Word Picture Vocabulary Test and measures of intelligence, receptive vocabulary, and visual-motor coordination in borderline and mildly retarded children. *Psychology in the Schools*, 19, 315-318.

Gottfied, A., Guerin, D., Spencer, J. & Meyer, C. (1983). Concurrent validity of the Minnesota Child Development Inventory in a nonclinical sample. *Journal of Consulting and Clinical Psychology*, 51, 643-644.

Gottfried, A., Guerin, D., Spencer, J. & Meyer, C. (1984) Validity of the Minnesota Child Development Inventory in screening young children's developmental status. *Journal of Pediatric Psychology*, 9, 219-229.

Greer, S., Buachner, H., & Zuckerman, B. (1989). The Denver Developmental Screening Test: How good is its predictive validity? *Developmental Medicine and Child Neurology*, 31, 774-781.

Guerin, D. & Gottfried, A. (1987). Minnesota Child Development Inventories: Predictors of intelligence, achievement, and adaptability. *Journal of Pediatric Psychology*, 12, 595-609.

Herman, D. (1994) Expressive One Word Picture Vocabulary Test Revised. In D. Keyser & R. Sweetland (Eds.), *Test Critiques* (Vol.10; pp. 240-243). Austin TX: Pro-Ed. Inc.

Illingworth, R. (1975). *The development of the infant and young child, normal and abnormal*. Edinburgh: Livingstone.

Ireton, H. & Thwing, E. (1974) *Manual for the Minnesota Child Development Inventory*. Minneapolis, MN: Behaviour Science Systems.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in spanish speaking infants and toddlers. *Journal of Child Language*, 20, 523-549.

Jacob, S., Snider, K., & Wilson, J. (1988). Validity of the DIAL-R for identifying children with special needs and predicting early reading achievement. *Journal of Psychoeducational Assessment*, 6, 289-297.

Kagan, S. (1992). Readiness past, present and future: Shaping the agenda. *Young Children*, November: 48-53.

Keith, T., Fehrman, P., Harrison, P., & Pottebaum, S. (1987). The relation between adaptive behaviour and intelligence: Testing alternative explanations. *Journal of School Psychology*, 25, 31-43.

Keith, T., Harrison, P. & Ehly, S. (1986) Effects of adaptive behavior on achievement: Path analysis of a national sample. cited in Harrison, P. Research with adaptive behavior scales.(1987) *Journal of Special Education*, 21, 37-68.

Kopparthi, R., McDermott, C., Lenke, M., Getz, M. & Frey, M. (1991). The Minnesota Child Development Inventory: Validity and reliability for assessing development in infancy. *Developmental and Behavioral Pediatrics*, 12, 217-222.

Kupersmidt, J., Coie, J. & Dodge, K. (1990). The role of poor peer relationships in the development of disorder. In S. Asher and J. Coie (Eds.)., *Peer rejection in childhood*. (P. 274-305).

Kutsick, K., Vance, B., Schwarting, G., & West, R. (1988) A comparison of three different measures of intelligence with preschool children identified at risk. *Psychology in the Schools*, 25, 270-275.

Ladd, G. (1990). Having friends, keeping friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development*, 61, 1081-1100.

Lindquist, G. (1982). Preschool screening as a means of predicting later reading achievement. Journal of Learning Disabilities, 15, 331-332.

Mardell, C. & Goldenberg, D. (1975) *DIAL Manual*. Edison, NJ: Childcraft Education Corporation.

Mardell-Czudnowski, C., Goldenberg, D., Suen, H., & Fries, R. (1988). Predictive validity of DIAL-R. *Diagnostique*, 14, 55-62.

McMann, G., & Barnett, D.(1984). An analysis of the construct validity of two measures of adaptive behaviour. *Journal of Psychoeducational Assessment*, 2, 239-247.

Meisels, S. (1984). Prediction, prevention, and developmental screening in the PSDT program. In H. Stevenson & A. Siegel (Eds), *Child development research and social policy* (pp.267-317). Chicago: University of Chicago Press.

Miller, L. & Sprong, T. (1986). Psychometric and qualitative comparison of four preschool screening instruments. *Journal of Learning Disabilities*, 19, 480-484.

Nugent, J. (1976). A comment on the efficacy of the Revised Denver Developmental Screening Test. *American Journal of Mental Deficiency*, 80, 570-572.

Obrzut, J., Bolocofsky, D., Heath, C., & Jones, M. (1981). An investigation of the DIAL as a prekindergarten screening instrument. *Educational and Psychological Measurement*, 41, 1231-1241.

O'Hanlon, L., Washkevich, D., & Thal, D, (1991). *MacArthur Communicative Development Inventory-Toddlers: validation for language impaired children.* Poster presented at the Annual Convention of the American Speech Language Hearing Association, Atlanta, Georgia. Platt, L., Kamphaus, R., Cole, R., & Smith, C. (1991). Relationship between adaptive behaviour and intelligence: Additional evidence. *Psychological Reports*, 68, 139-145.

Pulkkinen, L. & Tremblay, R. (1992). Patterns of boys' social adjustment in two cultures and at different ages: A longitudinal perspective. *International Journal of Behavioral Development*, 15, 527-553.

Ross, D. & Shillington, R. (1990). Child poverty and poor educational attainment: The economic costs and implications for society. In *Children in poverty: Toward a better future*. Report of the standing committee on Social Affairs, Science and Technology. Ottawa: Minister of Supply Services, Appendix I.

Sandford, M., Offord, D., McLeod, K., Boyle, M., Byrne., C. & Hall, B. (1994). Pathways into the workforce: Antecedents of school and work force status. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33, 1036-1046.

Sattler, J. (1992). *Assessment of Children* (3 ed.). San Diego, CA: Jerome Sattler Publisher Inc.

Saylor, C., & Brandt, B. (1986). The Minnesota Child Development Inventory: A valid maternal report form for assessing development in infancy. *Journal of Developmental and Behavioral Pediatrics*, 7, 308-311.

Schraeder, B. (1993). Assessment of measures to detect preschool academic risk in very-low-birthweight children. *Nursing Research*, 42, 17-21.

Shoemaker, O., Saylor, C., & Erickson, M. (1993). Concurrent validity of the Minnesota Child Development Inventory with High Risk Infants. *Journal of Pediatric Psychology*, 18, 377-388.

Snyder, P., Snyder, R., & Massong, S. (1981). The Visual Motor Integration Test: High interjudge reliability, high potential for diagnostic error. *Psychology in the Schools*, 18, 55-59.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). Vineland Adaptive Behavior Scales: Interview Edition, Survey Form manual. Circle Pines, MN: American Guidance Service.

Statistics Canada (1996). School leavers follow up survey, 1995. *The Daily*, October 16, 1996.

Suen, H., Mardell-Czudnowski, C., & Goldenberg, D. (1989). Classification reliability of the DIAL-R preschool screening test. *Educational and Psychological Measurement*, 49, 673-680.

Thomas, A., Chess, S., & Korn, S. (1982). The reality of difficult temperament. *Merrill-Palmer Quarterly*, 28, 1-20.

Tomasselo, M. & Mervis, C. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, 59, 174-179.

Tomblin, B., Shonrock, C., & Hardy, J. (1989). The concurrent validity of the Minnesota Child Development Inventory as a measure of young children's language development. *Journal of Speech and Hearing Disorders*, 54, 101-105.

Umberger, F. (1985). Peabody Picture Vocabulary Test-Revised. In D. Keyser & R. Sweetland (Eds.), Test Critiques (Vol. 10; pp. 488-495). Kansas City, MO: Test Corporation of America.

Vacc, N., Vacc, N. & Fogleman, M. (1987). Preschool screening: Using the DIAL as a predictor of first grade performance. *Journal of School Psychology*, 25, 45-51.

Vance, H.B., Fuller, G., & Lester, M. (1986). A comparison of the Minnesota Perceptual Diagnostic Test Revised and the Bender Gestalt. *Journal of Learning Disabilities*, 19, 211-214.

Vance, B., West, R., & Kutsick, K. (1989) Prediction of Wechsler Preschool and Primary Scale Intelligence IQ scores for preschool children using the Peabody Picture Vocabulary Test-R and the Expressive One Word Picture Vocabulary Test. *Journal of Clinical Psychology*, 45, 642-644.

Worobey, J. & Worobey, H. (1997, April) *Preschool breakfast evaluation study: Effects on attention and discrimination*. Paper presented at the biennial meetings of the Society for Research in Child Development, Washington, DC.

Zaichkowsky, L., Zaichkowsky, L. & Martinek, T. (1980). *Growth and development: The child and physical activity*. St. Louis: CV Mosby.